

The Feasibility of Scalable Video Streaming over Femtocell Networks

Donglin Hu and Shiwen Mao

Department of Electrical and Computer Engineering

Auburn University, Auburn, AL, USA

Abstract

In this paper, we consider femtocell CR networks, where femto base stations (FBS) are deployed to greatly improve network coverage and capacity. We investigate the problem of generic data multicast in femtocell networks. We reformulate the resulting MINLP problem into a simpler form, and derive upper and lower performance bounds. Then we consider three typical connection scenarios in the femtocell network, and develop optimal and near-optimal algorithms for the three scenarios. Second, we tackle the problem of streaming scalable videos in femtocell CR networks. A framework is developed to captures the key design issues and trade-offs with a stochastic programming problem formulation. In the case of a single FBS, we develop an optimum-achieving distributed algorithm, which is shown also optimal for the case of multiple non-interfering FBS's. In the case of interfering FBS's, we develop a greedy algorithm that can compute near-optimal solutions, and prove a closed-form lower bound on its performance.

I. INTRODUCTION

Due to the use of open space as transmission medium, capacity of wireless networks are usually limited by interference. When a mobile user moves away from the base station, a considerably larger transmit power is needed to overcome attenuation, while causing interference to other users and deteriorating network capacity. To this end, femtocells provide an effective solution that brings network infrastructure closer to mobile users. A femtocell is a small (e.g., residential) cellular network, with a *femto base station* (FBS) connected to the owner's broadband wireline network [1]–[3]. The FBS serves approved users when they are within the coverage. Among the many benefits, femtocells are shown effective on improving network coverage and capacity [1].

Due to reduced distance, transmit power can be greatly reduced, leading to prolonged battery life, improved signal-to-interference-plus-noise ratio (SINR), and better spatial reuse of spectrum.

Femtocells have received significant interest from the wireless industry. Although highly promising, many important problems should be addressed to fully harvest their potential, such as interference mitigation, resource allocation, synchronization, and QoS provisioning [1], [2]. It is also critical for the success of this technology to support important applications such as real-time video streaming in femtocell networks.

In this paper, we first investigate the problem of data multicast in femtocell networks. It is not atypical that many users may request for the same content, as often observed in wireline networks. By allowing multiple users to share the same downlink multicast transmission, significant spectrum and power savings can be achieved.

In particular, we adopt *superposition coding* (SC) and *successive interference cancellation* (SIC), two well-known PHY techniques, for data multicast in femtocell networks [4]. With SC, a compound signal is transmitted, consisting of multiple signals (or, layers) from different senders or from the same sender. With SIC, a strong signal can be first decoded, by treating all other signals as noise. Then the decoder will reconstruct the signal from the decoded bits, and subtract the reconstructed signal from the compound signal. The next signal will be decoded from the residual, by treating the remaining signals as noise. And so forth. A special strength of the SC with SIC approach is that it enables simultaneous unicast transmissions (e.g., many-to-one or one-to-many). It has been shown that SC with SIC is more efficient than PHY techniques with orthogonal channels [4], [5].

We adopt SC and SIC for the unique femtocell network environment, and investigate how to enable efficient data multicast from the femtocells to multiple users. We formulate a Mixed Integer Nonlinear Programming (MINLP) problem, which is NP-hard in general. The objective is to minimize the total BS power consumption. Then we reformulate the MINLP problem into a simpler form, and derive upper and lower performance bounds. We also derive a simple heuristic scheme that assigns users to the BS's with a greedy approach. Finally, we consider three typical connection scenarios in the femtocell network, and develop optimal and near-optimal algorithms for the three scenarios. The proposed algorithms have low computational complexity, and are shown to outperform the heuristic scheme with considerable gains.

Then, we investigate the problem of video streaming in femtocell cognitive radio (CR) net-

works. We consider a femtocell network consisting of a *macro base station* (MBS) and multiple FBS's. The femtocell network is co-located with a primary network with multiple licensed channels. This is a challenging problem due to the stringent QoS requirements of real-time videos and, on the other hand, the new dimensions of network dynamics (i.e., channel availability) and uncertainties (i.e., spectrum sensing and errors) found in CR networks.

We adopt Scalable Video Coding (SVC) in our system [6], [7]. SVC encodes a video into multiple substreams, subsets of which can be decoded to provide different quality levels for the reconstructed video [8]. Such scalability is very useful for video streaming systems, especially in CR networks, to accommodate heterogeneous channel availabilities and dynamic network conditions. We consider H.264/SVC medium grain scalable (MGS) videos, since MGS can achieve better rate-distortion performance over Fine-Granularity-Scalability (FGS), although it only has Network Abstraction Layer (NAL) unit-based granularity [8].

The unique femtocell network architecture and the scalable video allow us to develop a framework that captures the key design issues and trade-offs, and to formulate a *stochastic programming* problem. It has been shown that the deployment of femtocells has a significant impact on the network performance [1]. In this paper, we examine three deployment scenarios. In the case of a single FBS, we apply *dual decomposition* to develop a distributed algorithm that can compute the optimal solution. In the case of multiple non-interfering FBS's, we show that the same distributed algorithm can be used to compute optimal solutions. In the case of multiple interfering FBS's, we develop a greedy algorithm that can compute near-optimal solutions, and prove a closed-form lower bound for its performance based on an *interference graph* model. The proposed algorithms are evaluated with simulations, and are shown to outperform three alternative schemes with considerable gains.

The remainder of this paper is organized as follows. The related work is discussed in Section II. We investigate the problem of data multicast over femtocell networks in Section III. The problem of streaming multiple MGS videos in a femtocell CR network is discussed in Section IV. Section V concludes this paper.

II. BACKGROUND AND RELATED WORK

Femtocells have attracted considerable interest from both industry and academia. Technical and business challenges, requirements and some preliminary solutions to femtocell networks are

discussed in [1]. Since FBS's are distributedly located and are able to spatially reuse the same channel, considerable research efforts were made on interference analysis and mitigation [9], [10]. A distributed utility based SINR adaptation scheme was presented in [9] to alleviate cross-tire interference at the macrocell from co-channel femtocells. Lee, Oh and Lee [10] proposed a fractional frequency reuse scheme to mitigate inter-femtocell interference.

Deploying femtocells by underlaying the macrocell has been proved to significantly improve indoor coverage and system capacity. However, interference mitigation in a two-tier heterogeneous network is a challenging problem. In [11], the interference from macrocell and femtocells was mitigated by a spatial channel separation scheme with codeword-to-channel mapping. In [12], the rate distribution in the macrocell was improved by subband partitioning and modest gains were achieved by interference cancellation. In [13], the interference was controlled by denying the access of femtocell base stations to protect the transmission of nearby macro base station. A novel algorithmic framework was presented in [14] for dynamic interference management to deliver QoS, fairness and high system efficiency in LTE-A femtocell networks. Requiring no modification of existing macrocells, CR was shown to achieve considerable performance improvement when applied to interference mitigation [15]. In [16], the orthogonal time-frequency blocks and transmission opportunities were allocated based on a safe/victim classification.

SIC has high potential of sending or receiving multiple signals concurrently, which improves the transmission efficiency [17]. In [5], the authors developed MAC and routing protocols that exploit SC and SIC to enable simultaneous unicast transmissions. Sen, et al. investigated the possible throughput gains with SIC from a MAC layer perspective [18]. Power control for SIC was comprehensively investigated and widely applied to code division multiple access (CDMA) systems [19]–[23]. Applying game theory, Jean and Jabbari proposed an uplink power control under SIC in direct sequence-CDMA networks [19]. In [20], the authors introduced an iterative two-stage SIC detection scheme for a multicode MIMO system and showed the proposed scheme significantly outperformed the equal power allocation scheme. A scheme on joint power control and receiver optimization of CDMA transceivers was presented in [21]. In [22], [23], the impact of imperfect channel estimation and imperfect interference cancellation on the capacity of CDMA systems was examined.

The problem of video over CR networks has only been studied in a few recent papers [24]–[27]. In our prior work, we investigated the problem of scalable video over infrastructure-based

CR networks [6] and multi-hop CR networks [7]. The preliminary results of video over femtocell CR networks were presented in [28].

III. MULTICAST IN FEMTOCELL NETWORKS WITH SUPERPOSITION CODING AND SUCCESSIVE INTERFERENCE CANCELLATION

In this section, we formulate a Mixed Integer Nonlinear Programming (MINLP) problem of data multicast in femtocell networks, which is NP-hard in general. Then we reformulate the MINLP problem into a simpler form, and derive upper and lower performance bounds. We also derive a simple heuristic scheme that assigns users to the BS's with a greedy approach. Finally, we consider three typical connection scenarios in the femtocell network, and develop optimal and near-optimal algorithms for the three scenarios. The proposed algorithms have low computational complexity, and are shown to outperform the heuristic scheme with considerable gains.

A. System Model and Problem Statement

1) *System Model*: Consider a femtocell network with an MBS (indexed 0) and M FBS's (indexed from 1 to M) deployed in the area. The M FBS's are connected to the MBS and the Internet via broadband wireline connections. Furthermore, we assume a spectrum band that is divided into two parts, one is allocated to the MBS with bandwidth B_0 and the other is allocated to the M FBS's. The bandwidth allocated to FBS m is denoted by B_m . When there is no overlap between the coverages of two FBS's, they can spatially reuse the same spectrum. Otherwise, the MBS allocates disjoint spectrum to the FBS's with overlapping coverages. We assumed the spectrum allocation is known a priori.

There are K mobile users in the femtocell network. Each user is equipped with one transceiver that can be tuned to one of the two available channels, i.e., connecting to a nearby FBS or to the MBS. The network is time slotted. We assume block-fading channels, where the channel condition is constant in each time slot [4]. We focus on a multicast scenario, where the MBS and FBS's multicast a data file to the K users. The data file is divided into multiple packets with equal length and transmitted in sequence with the same modulation scheme. Once packet l is successfully received and decoded at the user, it requests packet $(l + 1)$ in the next time slot.

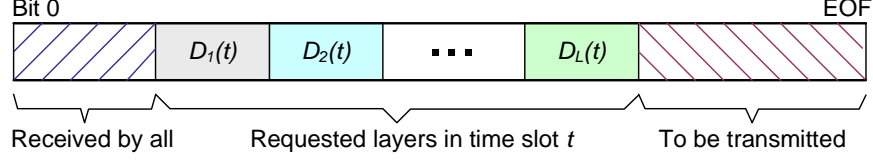


Fig. 1. Superposition coding and successive interference cancellation.

We adopt SC and SIC to transmit these packets [4], as illustrated in Fig. 1. In each time slot t , the compound signal has L layers (or, levels), denoted as $D_1(t), \dots, D_L(t)$. Each level $D_i(t)$, $i = 1, \dots, L$, is a packet requested by some of the users in time slot t . A user that has successfully decoded $D_i(t)$, for all $i = 1, \dots, l-1$, is able to subtract these signals from the received compound signal and then decodes $D_l(t)$, while the signals from $D_{l+1}(t)$ to $D_L(t)$ are treated as noise.

2) *Problem Statement:* For the SC and SIC scheme to work, the transmit powers for the levels should be carefully determined, such that there is a sufficiently high SNR for the levels to be decodable. It is also important to control the transmit powers of the BS's to reduce interference and leverage frequency reuse. The annual power bill is a large part of a mobile operator's costs [29]. Minimizing BS power consumption is important to reduce not only the operator's OPEX, but also the global CO₂ emission; an important step towards "green" communications.

Therefore, we focus on BS power allocation in this paper. The objective is to minimize the total power of all the BS's, while guaranteeing a target rate R_{tar} for each user. Recall that the data file is partitioned into equal-length packets. The target rate R_{tar} ensures that a packet can be transmitted within a time slot, for given modulation and channel coding schemes.

Define binary indicator I_m^k , for all m and k , as:

$$I_m^k = \begin{cases} 1, & \text{if user } k \text{ connects to BS } m \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Consider a general time slot t when L data packets, or levels, are requested. We formulate the

optimal power allocation problem (termed OPT-Power) as follows.

$$\text{minimize: } \sum_{m=0}^M \sum_{l=1}^L P_l^m \quad (2)$$

$$\text{subject to: } B_m \log_2(1 + \gamma_m^k I_m^k) \geq R_{tar} I_m^k, \text{ for all } k \quad (3)$$

$$\sum_{m=0}^M I_m^k = 1, \text{ for all } k \quad (4)$$

$$P_l^m \geq 0, \text{ for all } l, m, \quad (5)$$

where P_l^m is the power of BS m for transmitting the level l packet; γ_m^k is the SNR at user k if it connects to BS m . Constraint (3) guarantees the minimum rate at each user. Constraint (4) is due to the fact that each user is equipped with one transceiver, so it can only connect to one BS.

Let \mathcal{U}_l denote the set of users requesting the level l packet. A user $k \in \mathcal{U}_l$ has decoded all the packets up to D_{l-1} . It subtracts the decoded signals from the received signal and treats signals D_{l+1}, \dots, D_L as noise. The SNR at user $k \in \mathcal{U}_l$, for $l = 1, \dots, L-1$, can be written as:

$$\gamma_m^k = H_m^k P_l^m / \left(N_0 + H_m^k \sum_{i=l+1}^L P_i^m \right), \quad (6)$$

where H_m^k is the random channel gain from BS m to user k and N_0 is the noise power. For user $k \in \mathcal{U}_L$ that requests the last packet D_L , the SNR is

$$\gamma_m^k = H_m^k P_L^m / N_0. \quad (7)$$

The optimization variables in Problem OPT-Power consist of the binary variables I_m^k 's and the continuous variables P_l^m 's. It is an MINLP problem, which is NP-hard in general. In Section III-B, we first reformulate the problem to obtain a simpler form, and then develop effective algorithms for optimal and suboptimal solutions.

B. Reformulation and Power Allocation

In this section, we reformulate Problem OPT-Power to obtain a simpler form, and derive an upper bound and a lower bound for the total BS power. The reformulation also leads to a simple heuristic algorithm. Finally, we introduce power allocation algorithms for three connection scenarios.

1) *Problem Reformulation:* Due to the monotonic logarithm functions and the binary indicators I_m^k , constraint (3) can be rewritten as:

$$\gamma_m^k I_m^k \geq \Gamma_m^k I_m^k, \quad m = 0, 1, \dots, M, \quad (8)$$

where $\Gamma_m^k = \Gamma_m =: 2^{R_{tar}/B_m} - 1$ is the minimum SNR requirement at user k that connects to BS m . To further simplify the problem, define $Q_l^m = \sum_{i=l}^L P_i^m$, with $Q_{L+1}^m = 0$. Then power P_l^m is the difference

$$P_l^m = Q_l^m - Q_{l+1}^m. \quad (9)$$

Problem OPT-Power can be reformulated as:

$$\text{minimize } \sum_{m=0}^M Q_1^m \quad (10)$$

$$\text{subject to: } H_m^k (Q_l^m - Q_{l+1}^m) / (N_0 + H_m^k Q_{l+1}^m) I_m^k \geq \Gamma_m I_m^k, \quad \text{for all } k \in \mathcal{U}_l, l = 1, \dots, L \quad (11)$$

$$Q_l^m \geq Q_{l+1}^m, l = 1, \dots, L \quad (12)$$

$$\sum_{m=0}^M I_m^k = 1, \text{ for all } k. \quad (13)$$

For $l \leq L$, constraint (11) can be rewritten as:

$$Q_l^m I_m^k \geq [N_0 \Gamma_m / H_m^k + (1 + \Gamma_m) Q_{l+1}^m] I_m^k. \quad (14)$$

Let \mathcal{U}_l^m be the subset of users connecting to BS m in \mathcal{U}_l . Since $Q_l^m \geq Q_{l+1}^m$, (14) can be rewritten as,

$$Q_l^m = \max \left\{ Q_{l+1}^m, \max_{k \in \mathcal{U}_l^m} [N_0 \Gamma_m / H_m^k + (1 + \Gamma_m) Q_{l+1}^m] \right\}. \quad (15)$$

From (15), we define a function $Q_l^m = F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ as:

$$F_m(Q_{l+1}^m, \mathcal{U}_l^m) = \begin{cases} Q_{l+1}^m, & \mathcal{U}_l^m = \emptyset \\ \max_{k \in \mathcal{U}_l^m} \left\{ \frac{N_0 \Gamma_m}{H_m^k} + (1 + \Gamma_m) Q_{l+1}^m \right\}, & \mathcal{U}_l^m \neq \emptyset. \end{cases} \quad (16)$$

Obviously, $F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ is non-decreasing with respect to Q_{l+1}^m . It follows that

$$\begin{aligned} Q_1^m &= F_m(Q_2^m, \mathcal{U}_1^m) = F_m(F_m(Q_3^m, \mathcal{U}_2^m), \mathcal{U}_1^m) \\ &= F_m(\dots (F_m(Q_{L+1}^m, \mathcal{U}_L^m), \mathcal{U}_{L-1}^m), \dots, \mathcal{U}_1^m) \\ &= F_m(\dots (F_m(0, \mathcal{U}_L^m), \mathcal{U}_{L-1}^m), \dots, \mathcal{U}_1^m). \end{aligned} \quad (17)$$

If none of the subsets \mathcal{U}_l^m ($l = 1, \dots, L$) is empty, we can expand the above recursive term using (16). It follows that

$$Q_1^m = N_0 \Gamma_m \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \{1/H_m^k\}, \quad (18)$$

where the exponent c_l^m is defined as $c_1^m = 0$ and $c_{l+1}^m = c_l^m + 1$. Otherwise, if a subset $\mathcal{U}_l^m = \emptyset$ for some m , we have that $Q_l^m = Q_{l+1}^m$, $\max_{k \in \mathcal{U}_l^m} \{1/H_m^k\} = \max_{k \in \emptyset} \{1/H_m^k\} = 0$, and $c_l^m = c_{l-1}^m$. Eq. (18) still holds true.

Finally, the objective function (10) can be rewritten as

$$\sum_{m=0}^M N_0 \Gamma_m \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \{1/H_m^k\}. \quad (19)$$

Since $(1 + \Gamma_m) > 0$, it can be seen that to minimize the total BS power, we need to keep the c_l^m 's as low as possible.

2) *Performance Bounds* : The reformulation and simplification allow us to derive performance bounds for the total BS power consumption. First, we derive the upper bound for the objective function (10). Define a variable

$$\overline{G}_m = \max_{l \in \{1, \dots, L\}} \max_{k \in \mathcal{U}_l^m} \{\Gamma_m/H_m^k\}, \quad (20)$$

which corresponds to the user with the worst channel condition among all users that connect to BS m . It follows that:

$$\begin{aligned} \sum_{m=0}^M Q_1^m &= N_0 \sum_{m=0}^M \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \{\Gamma_m/H_m^k\} \\ &\leq N_0 \sum_{m=0}^M \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \overline{G}_m \\ &\leq N_0 \sum_{m=0}^M \overline{G}_m \sum_{l=1}^L (1 + \Gamma_m)^{l-1} \\ &= N_0 \sum_{m=0}^M \overline{G}_m [(1 + \Gamma_m)^L - 1] / \Gamma_m. \end{aligned} \quad (21)$$

In (21), the first inequality is from the definition of \overline{G}_m . The second inequality is from the definition of c_l^m . Specifically, $c_1^m = 0$; when $\mathcal{U}_l^m \neq \emptyset$, we have $c_l^m = c_{l-1}^m + 1$; when $\mathcal{U}_l^m = \emptyset$, we have $c_l^m = c_{l-1}^m$. It follows that $c_l^m \leq l - 1$. Therefore, (21) is an upper bound on the objective function (10).

Furthermore, by defining $\overline{G} = \max_{m \in \{0, \dots, M\}} \{\overline{G}_m\}$, and $\overline{\Gamma} = \max_{m \in \{0, \dots, M\}} \{\Gamma_m\}$, we can get a looser upper bound from 21 as

$$\sum_{m=0}^M Q_1^m \leq N_0 \overline{G} (M+1) [(1 + \overline{\Gamma})^L - 1] / \overline{\Gamma}. \quad (22)$$

Next, we derive a lower bound for (10). Define

$$\begin{cases} \underline{G}^l = \min_{m \in \{0, \dots, M\}} \max_{k \in \mathcal{U}_l^m} \{\Gamma_m / H_m^k\} \\ \underline{\Gamma} = \min_{m \in \{0, \dots, M\}} \{\Gamma_m\}. \end{cases} \quad (23)$$

We have that

$$\begin{aligned} \sum_{m=0}^M Q_1^m &= N_0 \sum_{m=0}^M \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \max_{k \in \mathcal{U}_l^m} \{\Gamma_m / H_m^k\} \\ &\geq N_0 \sum_{m=0}^M \sum_{l=1}^L (1 + \Gamma_m)^{c_l^m} \underline{G}^l \\ &\geq N_0 \sum_{l=1}^L \underline{G}^l \sum_{m=0}^M (1 + \underline{\Gamma})^{c_l^m} \\ &\geq N_0 (M+1) \sum_{l=1}^L \underline{G}^l (1 + \underline{\Gamma})^{\frac{\sum_{m=0}^M c_l^m}{M+1}} \\ &\geq N_0 (M+1) \sum_{l=1}^L \underline{G}^l (1 + \underline{\Gamma})^{\frac{l-1}{M+1}}. \end{aligned} \quad (24)$$

In (24), the first inequality is from the definition of \underline{G}^l . The second inequality is due to the definition of $\underline{\Gamma}$. The third inequality is due to the fact that $(1 + \Gamma)^{c_l^m}$ is a convex function. The fourth inequality is because that each level must be transmitted by at least one BS. Thus for each level l , there is at least one $c_l^m = c_{l-1}^m + 1$ for some m . It follows that the sum $\sum_{m=0}^M c_l^m$ should be greater than $l - 1$. Therefore, (24) provides a lower bound for (10).

Furthermore, by defining $\underline{G} = \min_{l \in \{1, \dots, L\}} \{\underline{G}^l\}$, we can obtain a looser lower bound from (24) as

$$\sum_{m=0}^M Q_1^m \geq N_0 \underline{G} (M+1) \frac{(1 + \underline{\Gamma})^{\frac{L}{M+1}} - 1}{(1 + \underline{\Gamma})^{\frac{1}{M+1}} - 1}. \quad (25)$$

3) A Simple Heuristic Scheme : We first describe a greedy heuristic algorithm that solves OPT-Power with suboptimal solutions. With this heuristic, each user compares the channel gains from the MBS and the FBS's. It chooses the BS with the best channel condition to connect to, thus the values of the binary variables I_m^k are determined. Once the binary variables are fixed,

all the subsets \mathcal{U}_l^m 's are determined. Starting with $Q_{L+1}^m = 0$, we can apply (15) iteratively to find the Q_l^m 's. Finally, the transmit powers P_l^m can be computed using (9).

With this approach, among the users requesting the level l packet, it is more likely that some of them connect to the MBS and the rest connect to some FBS's, due to the random channel gains in each time slot. In this situation, both MBS and FBS will have to transmit all the requested data packets. Such situation is not optimal for minimizing the total power, as will be discussed in Section III-B4b.

4) *Power Allocation Algorithms* : In the following, we develop three power allocation algorithms for three different connection scenarios with a more structured approach.

a) *Case I–One Base Station*: We first consider the simplest connection scenario where all the K users connect to the same BS (i.e., either the MBS or an FBS). Assume all the users connect to BS m . Then we have $I_m^k = 1$ for all k , and all the subsets \mathcal{U}_l^m are non-empty; $I_{m'}^k = 0$ for all k and all $m' \neq m$, and all the subsets $\mathcal{U}_l^{m'}$ are empty for $m' \neq m$.

From (16), we can derive the optimal solution as:

$$\begin{aligned} Q_l^{m*} &= (1 + \Gamma_m)Q_{l+1}^{m*} + \max_{k \in \mathcal{U}_l^m} \{N_0\Gamma_m/H_m^k\}, \\ &= N_0\Gamma_m \sum_{i=l}^L (1 + \Gamma_m)^{i-l} \max_{k \in \mathcal{U}_i^m} \{1/H_m^k\}, \quad l = 1, 2, \dots, L. \end{aligned} \quad (26)$$

Recall that $Q_{L+1}^{m*} = Q_{L+1}^m = 0$, the optimal power allocation for Problem OPT-Power in this case is:

$$P_l^{m'*} = \begin{cases} Q_l^{m*} - Q_{l+1}^{m*}, & m' = m, \text{ for all } l \\ 0, & m' \neq m, \text{ for all } l. \end{cases} \quad (27)$$

b) *Case II–MBS and One FBS* : We next consider the case with one MBS and one FBS (i.e., $M = 1$), where each user has two choices: connecting to either the FBS or the MBS.

Recall that \mathcal{U}_l^0 and \mathcal{U}_l^1 are the subset of users who connected to the MBS and the FBS, respectively, and who request the level l packet. Examining (18), we find that the total power of BS m can be significantly reduced if one or more levels are not transmitted, since the exponent c_l^m will not be increased in this case. Furthermore, consider the two choices: (i) not transmitting level l , and (ii) not transmitting level $l' > l$ from BS m . The first choice will yield larger power savings, since more exponents (i.e., $c_l^m, c_{l+1}^m, \dots, c_{l'-1}^m$) will assume smaller values. Therefore, we should let these two subsets be empty whenever possible, i.e., either $\mathcal{U}_l^0 = \emptyset$ or $\mathcal{U}_l^1 = \emptyset$.

TABLE I
POWER ALLOCATION ALGORITHM FOR CASE II

1:	Initialize all c_l^0, c_l^1, Q_{L+1}^0 and Q_{L+1}^1 to zero;
2:	FOR $l = 1$ TO L
3:	IF $(\Gamma_0(1 + \Gamma_0)^{c_l^0} G_l^0 \leq \Gamma_1(1 + \Gamma_1)^{c_l^1} G_l^0)$
4:	Set $\mathcal{U}_l^0 = \mathcal{U}_l$ and $\mathcal{U}_l^1 = \emptyset$;
5:	$c_l^0 = c_l^0 + 1$;
6:	ELSE
7:	Set $\mathcal{U}_l^0 = \emptyset$ and $\mathcal{U}_l^1 = \mathcal{U}_l$;
8:	$c_l^1 = c_l^1 + 1$;
9:	END IF
10:	END FOR
11:	FOR $l = L$ TO 1
12:	$Q_l^0 = F_0(Q_{l+1}^0, \mathcal{U}_l^0)$ and $P_l^0 = Q_l^0 - Q_{l+1}^0$;
13:	$Q_l^1 = F_1(Q_{l+1}^1, \mathcal{U}_l^1)$ and $P_l^1 = Q_l^1 - Q_{l+1}^1$;
14:	END FOR

According to this policy, all the users requesting the level l packet will connect to the same BS. We only need to make the optimal connection decision for each subset of users requesting the same level of packet, rather than for each individual user.

Since not transmitting a lower level packet yields more power savings for a BS, we calculate the power from the lowest to the highest level, and decide whether connecting to the MBS or the FBS for users in each level. Define $G_l^0 = \max_{k \in \mathcal{U}_l} \{1/H_0^k\}$ and $G_l^1 = \max_{k \in \mathcal{U}_l} \{1/H_1^k\}$. The algorithm for solving Problem OPT-Power in this case is given in Table I. In Steps 2–10, the decision on whether connecting to the MBS or the FBS is made by comparing the expected increments in the total power. The user subsets \mathcal{U}_l^0 and \mathcal{U}_l^1 are determined in Steps 4 and 7. In Steps 11–14, Q_l^m 's and the corresponding P_l^m 's are computed in the reverse order, based on the determined subsets \mathcal{U}_l^0 and \mathcal{U}_l^1 .

The computational complexity of this algorithm is $\mathcal{O}(L)$.

c) Case III–MBS and Multiple FBS's : Finally, we consider the general case with one MBS and multiple FBS's in the network. Each user is able to connect to the MBS or a nearby FBS. Recall that we define \mathcal{U}_l as the set of users requesting the level l packet, and \mathcal{U}_l^m as the subset

of users in \mathcal{U}_l that *connect* to BS m . These sets have the following properties.

$$\begin{cases} \bigcup_{m=0}^M \mathcal{U}_l^m = \mathcal{U}_l \\ \mathcal{U}_l^m \cap \mathcal{U}_l^{m'} = \emptyset, \text{ for all } m' \neq m. \end{cases}$$

The first property is due to the fact that each user must connect to the MBS or an FBS. The second property is because each user can connect to only one BS. The user subsets connecting to different BS's do not overlap. Therefore, \mathcal{U}_l^m 's is a *partition* of \mathcal{U}_l with respect to m .

In addition, we define \mathcal{S}_l^m as the set of possible users that are *covered* by BS m and request the level l packet. These sets have the following properties.

$$\begin{cases} \bigcup_{m=1}^M \mathcal{S}_l^m = \mathcal{S}_l^0 = \mathcal{U}_l \\ \mathcal{S}_l^m \cap \mathcal{S}_l^0 = \mathcal{S}_l^m, \text{ for all } m \neq 0 \\ \mathcal{S}_l^m \cap \mathcal{S}_l^{m'} = \emptyset, \text{ for all } m' \neq m \text{ and } m, m' \neq 0. \end{cases}$$

The first property is because all users in each femtocell are covered by the MBS. The second property indicates that the users covered by FBS m are a subset of the users covered by the MBS. The third property shows that the user subsets in different femtocells do not overlap. We can see that the \mathcal{S}_l^m 's, for $m = 1, \dots, M$, are also a partition of \mathcal{U}_l .

Define $W_m(\mathcal{U}) = \max_{k \in \mathcal{U}} \{1/H_m^k\}$, where \mathcal{U} is the set of users and $m = 0, \dots, M$. If the set \mathcal{U} is empty, we define $W_m(\emptyset) = 0$. For example, consider Case II where $M = 1$. We have $\mathcal{S}_l^0 = \mathcal{S}_l^1 = \mathcal{U}_l$, $W_0(\mathcal{U}_l) = G_l^0$, and $W_1(\mathcal{U}_l) = G_l^1$.

The power allocation algorithm for Case III is presented in Table II. The algorithm iteratively picks users from the *eligible* subset \mathcal{S}_l^m and assigns them to the *allocated* subset \mathcal{U}_l^m . In each step l , Ψ is the subset of FBS's that will transmit the level l packet; the complementary set $\bar{\Psi}$ is the subset of FBS's that will not transmit the level l packet. The expected increment in total power for each partition is computed, and the partition with the smallest expected increment will be chosen. Δ_l^m is the power of BS m for transmitting the level l data packet. In Steps 6–15, the MBS and FBS combination Ψ is determined for transmitting the level l packet, with the lowest power Δ_0 . In Steps 16–30, elements in \mathcal{S}_l^m are assigned to \mathcal{U}_l^m according to Ψ . In Steps 31–35, power sums Q_l^m and the corresponding power allocations P_l^m are calculated in the reverse order from the known \mathcal{U}_l^m 's.

The complexity of the algorithm is $\mathcal{O}(ML)$.

TABLE II
POWER ALLOCATION ALGORITHM FOR CASE III

1:	Initialize: $c_l^m = 0$ and $Q_{L+1}^m = 0$, for all l, m ;
2:	FOR $l = 1$ TO L
3:	FOR $m = 0$ TO M
4:	$\Delta_l^m = \Gamma_m(1 + \Gamma_m)^{c_l^m} W_m(\mathcal{S}_l^m)$;
5:	END FOR
6:	Set $\Omega = \{1, \dots, M\}$ and $\Psi = \emptyset$;
7:	WHILE ($\Omega \neq \emptyset$)
8:	$m' = \arg \min_{m \in \Omega} \Delta_l^m$;
9:	Compute $\Delta' = \Gamma_0(1 + \Gamma_0)^{c_l^0} W_0(\bigcup_{m \in \Psi \cup m'} \mathcal{S}_l^m)$;
10:	IF ($(\sum_{m \in \Psi \cup m'} \Delta_l^m + \Delta') < \Delta_0$)
11:	Add m' to Ψ ;
12:	$\Delta_0 = \sum_{m \in \Psi} \Delta_l^m + \Delta'$;
13:	END IF
14:	Remove m' from Ω ;
15:	END WHILE
16:	IF ($\Psi = \emptyset$)
17:	$\mathcal{U}_l^0 = \mathcal{S}_l^0$;
18:	$c_l^0 = c_l^0 + 1$;
19:	Set $\mathcal{U}_l^m = \emptyset$, for all $m \neq 0$;
20:	ELSE
21:	$\mathcal{U}_l^0 = \bigcup_{m \in \Psi} \mathcal{S}_l^m$;
22:	IF ($ \Psi < M$)
23:	$c_l^0 = c_l^0 + 1$;
24:	END IF
25:	FOR $m \in \Psi$
26:	$c_l^m = c_l^m + 1$;
27:	$\mathcal{U}_l^m = \mathcal{S}_l^m$;
28:	END FOR
29:	END IF
30:	END FOR
31:	FOR $l = L$ TO 1
32:	FOR $m = 0$ TO M
33:	$Q_l^m = F_m(Q_{l+1}^m, \mathcal{U}_l^m)$ and $P_l^m = Q_l^m - Q_{l+1}^m$;
34:	END FOR
35:	END FOR

C. Performance Evaluation

We evaluate the performance of the proposed power allocation algorithms using MATLABTM. Three scenarios corresponding to the three cases in Section III-B are simulated: (i) Case I: a single MBS; (ii) Case II: one MBS and one FBS; and (iii) Case III: one MBS and three FBS's.

Since we do not find any similar schemes in the literature, we made the following comparisons. First, we compare Cases I and II with respect to BS power consumption and interference footprint. In both cases, there are $K = 8$ users and $L = 4$ levels. In Case I, the MBS bandwidth is $B_0 = 2$ MHz. In Case II, the MBS and the FBS share the 2 MHz total bandwidth; the MBS bandwidth is $B_0 = 1$ MHz and the FBS bandwidth is $B_1 = 1$ MHz. The target data rate R_{tar} is set to 2 Mbps. The channel gain from a base station to each user is exponentially distributed in each time slot.

The interference footprints in the three dimensional space are plotted in Fig. 2. The height B of the cylinders indicates the spectrum used by a BS, while the radius r is proportional to the BS transmit power. In Case I when only the MBS is used, the total BS power is 45.71 dBm and the volume of the cylinder is $\pi r^2 B = 18,841$ MHz m². In Case II when both the MBS and FBS are used, the total BS power is 34.58 dBm and the total volume of the two cylinders is 2,378 MHz m². Using an additional FBS achieves a 11.13 dB power saving and the interference footprint is reduced to 12.62% of that in Case I. This simple comparison clearly demonstrate the advantages of femtocells achieved by bringing BS's closer to users.

We next consider the more general Case III, using a femtocell network of one MBS and three FBS's. The MBS bandwidth is $B_0 = 1$ MHz and each FBS is assigned with bandwidth $B_m = 1$ MHz, $m = 1, 2, 3$. The target data rate is still 2 Mbps. In Figs. 3 and 4, we plot four curves, each obtained with: (i) the heuristic scheme described in Section III-B3; (ii) The proposed algorithm presented in Section III-B4c; (iii) The upper bound; and (iv) the lower bound derived in Section III-B2. Each point in the figures is the average of 10 simulation runs. The 95% confidence intervals are plotted as error bars, which are all negligible.

In Fig. 3, we examine the impact of the number of packet levels L on the total BS transmit power. We increase L from 2 to 6, and plot the total power of base stations. As expected, the more packet levels, the larger the BS power consumption. Both the proposed and heuristic curves lie in between the upper and lower bound curves. When L is increased from 2 to 6, the power

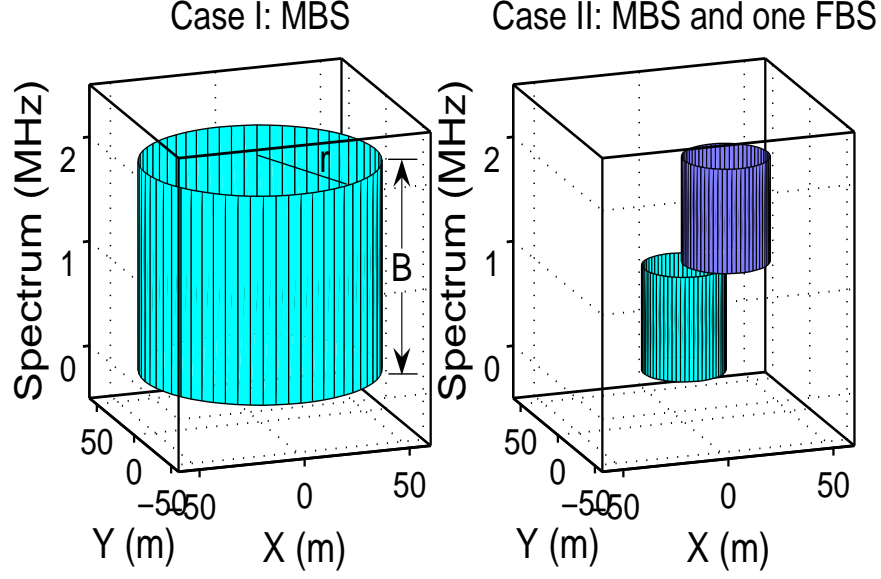


Fig. 2. Case I vs. Case II: interference footprints.

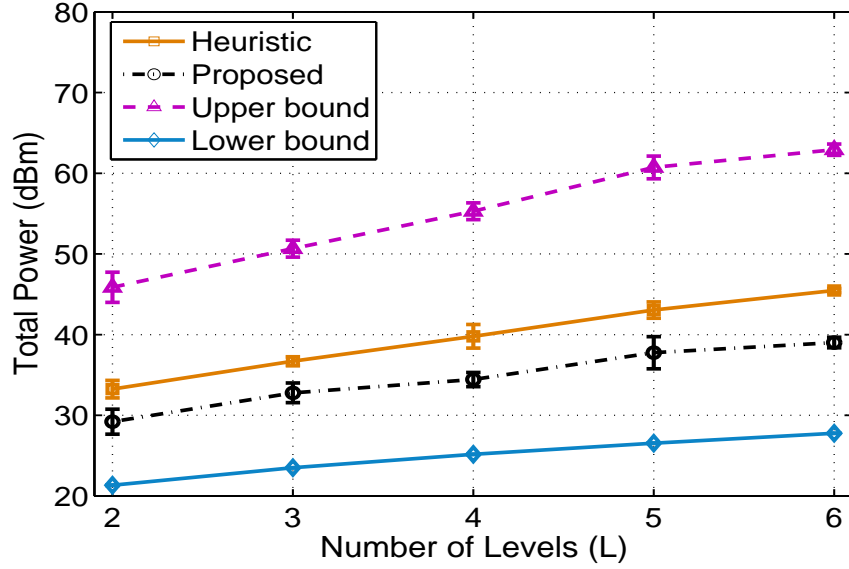


Fig. 3. Case III: impact of number of levels L .

consumption of the heuristic scheme is increased by 12.22 dB, while the power consumption of the proposed algorithm is increased by 9.94 dB. The power savings achieved by the proposed algorithm over the heuristic scheme range from 3.92 dB to 6.45 dB.

In Fig. 4, we show the impact of the BS bandwidths. The number of levels is $L = 4$. We fix

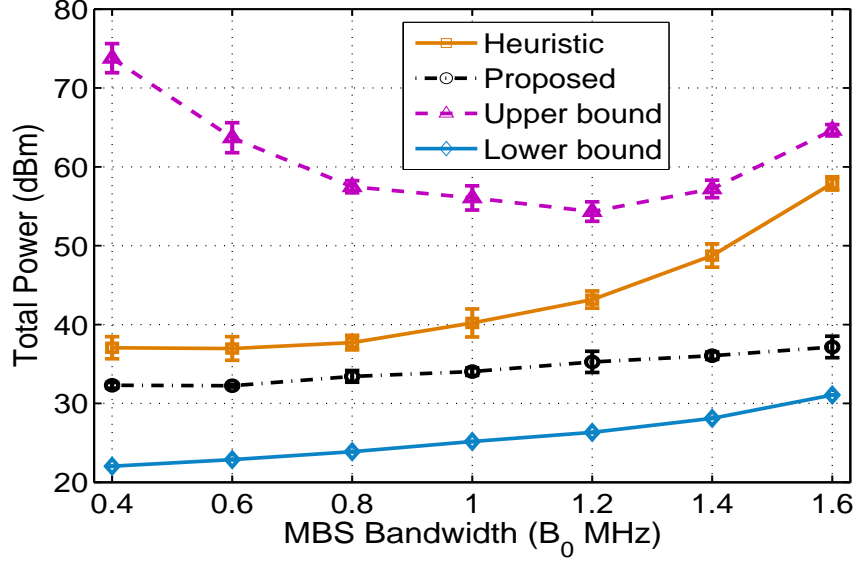


Fig. 4. Case III: impact of MBS bandwidth B_0 .

the total bandwidth at 2 MHz, which is shared by the MBS and FBS's. We increase the MBS bandwidth from 0.4 MHz to 1.6 MHz in steps of 0.2 MHz, while decrease the bandwidth of FBS's from 1.6 MHz to 0.4 MHz. We find that the total power consumption is increased as B_0 gets large. This is due to the fact that as the FBS bandwidth gets smaller, the FBS's have to spend more power to meet the minimum data rate requirement. The curve produced by the proposed algorithm has a smaller slope than that of the heuristic scheme: the overall increase in the total power of the proposed algorithm is 4.86 dB, while that of the heuristic scheme is 20.84 dB. This implies that the proposed scheme is not very sensitive to the bandwidth allocation between the MBS and FBS's. The proposed algorithm also achieves consider power savings over the heuristic scheme. When $B_0 = 1.6$ MHz, the total power of the proposed algorithm is 20.75 dB lower than that of the heuristic scheme.

IV. VIDEO OVER CR FEMTOCELL NETWORKS

In this section, we investigate the problem of video streaming in femtocell cognitive radio (CR) networks and formulate a *stochastic programming* problem to examine three deployment scenarios. In the case of a single FBS, we apply *dual decomposition* to develop an optimum-achieving distributed algorithm, which is shown also optimal for the case of multiple non-

interfering FBS's. In the case of multiple interfering FBS's, we develop a greedy algorithm that can compute near-optimal solutions, and prove a closed-form lower bound for its performance based on an *interference graph* model. The proposed algorithms are evaluated with simulations, and are shown to outperform three alternative schemes with considerable gains.

A. System Model and Preliminaries

1) *Spectrum and Network Model:* We consider a spectrum consisting of $(M + 1)$ channels, including one common, unlicensed channel (indexed as channel 0) and M licensed channels (indexed as channels 1 to M). The M licensed channels are allocated to a primary network, and the common channel is exclusively used by all CR users. We assume all the channels follow a synchronized time slot structure [30]. The capacity of each licensed channel is B_1 Mbps, while the capacity of the common channel is B_0 Mbps. The channel states evolve independently, while the occupancy of each licensed channel follows a two-state discrete-time Markov process.

The femtocell CR network is illustrated in Fig. 5. There is an MBS and N FBS's deployed in the area to serve CR users. The N FBS's are connected to the MBS (and the Internet) via broadband wireline connections. Due to advances in antenna technology, it is possible to equip multiple antennas at the base stations. The MBS has one antenna that is always tuned to the common channel. Each FBS is equipped with multiple antennas (e.g., M) and is able to sense multiple licensed channels at the beginning of each time slot. There are K_i CR users in femtocell i , $i = 1, 2, \dots, N$, and $\sum_{i=1}^N K_i = K$. Each CR user has a software radio transceiver, which can be tuned to any of the $M+1$ channels. A CR user will either connect to a nearby FBS using one or more of the licensed channels or to the MBS via the common channel.

Although the CR users are mobile, we assume constant topology during a time slot. If the topology is changed during a time slot, the video transmission will only be interrupted for the time slot, since the proposed algorithms are executed in every time slot for new channel assignment and schedule.

2) *Spectrum Sensing and Access:* The femtocell CR network is within the coverage of the infrastructure-based primary network. Both FBS's and CR users sense the channels to identify spectrum opportunities in each time slot. Each time slot consists of (i) a *sensing phase*, when CR users and FBS's sense licensed channels, (ii) a *transmission phase*, when CR users and FBS's

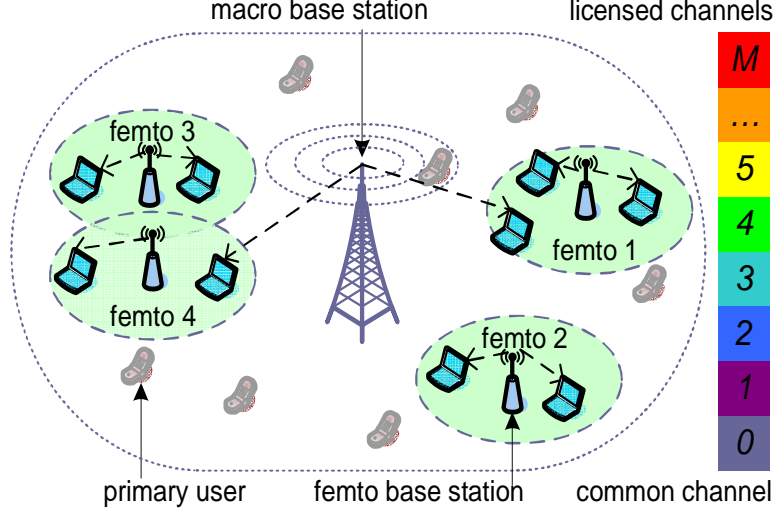


Fig. 5. A femtocell CR network with one MBS and four FBS's.

attempt to access licensed channels, and (iii) an *acknowledgment phase*, when acknowledgments (ACK) are returned to the source.

Cooperative sensing policy is also adopted here. We also adopt a *hypothesis test* to detect channel availability. We assume that each CR user chooses one channel to sense in each time slot, since it only has one transceiver. The sensing results will be shared among CR users and FBS's via the common channel in the sensing phase. Given L sensing results on channel m , i.e., $P_m^A(\Theta_1^m, \dots, \Theta_L^m)$, can be computed iteratively as follows.

$$P_m^A(\Theta_1^m) = \left[1 + \frac{\eta_m}{1 - \eta_m} \times \frac{(\delta_1^m)^{1-\Theta_1^m} (1 - \delta_1^m)^{\Theta_1^m}}{(\epsilon_1^m)^{\Theta_1^m} (1 - \epsilon_1^m)^{1-\Theta_1^m}} \right]^{-1} \quad (28)$$

$$\begin{aligned} P_m^A(\vec{\Theta}_l^m) &= P_m^A(\Theta_1^m, \Theta_2^m, \dots, \Theta_l^m) \\ &= \left\{ 1 + \left[\frac{1}{P_m^A(\Theta_1^m, \Theta_2^m, \dots, \Theta_{l-1}^m)} - 1 \right] \times \right. \\ &\quad \left. \frac{(\delta_l^m)^{1-\Theta_l^m} (1 - \delta_l^m)^{\Theta_l^m}}{(\epsilon_l^m)^{\Theta_l^m} (1 - \epsilon_l^m)^{1-\Theta_l^m}} \right\}^{-1}, l = 2, \dots, L. \end{aligned} \quad (29)$$

We adopt a probabilistic approach: based on sensing results $\vec{\Theta}_m$, we have $D_m(t) = 0$ with probability $P_m^D(\vec{\Theta}_m)$ and $D_m(t) = 1$ with probability $1 - P_m^D(\vec{\Theta}_m)$. For primary user protection, the collision probability with primary users caused by CR users should be bounded. The probability $P_m^D(\vec{\Theta}_m)$ is determined as follows

$$P_m^D(\vec{\Theta}_m) = \min \left\{ \gamma_m / \left[1 - P_m^A(\vec{\Theta}_m) \right], 1 \right\}. \quad (30)$$

Let $\mathcal{A}(t) := \{m | D_m(t) = 0\}$ be the set of available channels in time slot t . Then $G^t = \sum_{m \in \mathcal{A}(t)} P_m^A(\Theta_1^m)$ is the expected number of available channels. These channels will be accessed in the transmission phase of time slot t .

3) *Channel Model*: Without loss of generality, we consider independent block fading channels that is widely used in prior work [31]. The channel fading-gain process is piecewise constant on blocks of one time slot, and fading in different time slots are independent. Let $f_X^{i,j}(x)$ denote the *probability density function* of the received SINR X from a base station i at CR user j . We assume the packet can be successfully decoded if the received SINR exceeds a threshold H . The packet loss probability from base station i to CR user j is

$$P_{i,j} = \Pr\{X \leq H\} = \int_0^H f_X^{i,j}(x) dx = F_X^{i,j}(H), \quad (31)$$

where $F_X^{i,j}(H)$ is the cumulative density function of X .

In the case of correlated fading channels, which can be modeled as finite state Markov Process [32], the packet loss probability in the next time slot can be estimated from the known state of the previous time slot and the transition probabilities. If the packet is successfully decoded, the CR user returns an ACK to the base station in the ACK phase. We assume ACKs are always successfully delivered.

4) *Video Performance Measure*: We assume each active CR user receives a real-time video stream from either the MSB or an FSB. Without loss of generality, we adopt the MGS option of H.264/SVC, for scalability to accommodate the high variability of network bandwidth in CR networks.

Due to real-time constraint, each Group of Pictures (GOP) of a video stream must be delivered in the next T time slots. With MGS, enhancement layer NAL units can be discarded from a quality scalable bit stream, and thus packet-based quality scalable coding is provided. Our approach is to encode the video according to the maximum rate the channels can support. During transmission, only part of the MGS video gets transmitted as allowed by the current available channel bandwidth. The video packets are transmitted in decreasing order of their significance in decoding. When a truncated MGS video is received and decoded, the PSNR is computed by substituting the effective rate of the received MGS video into (32) given below, thus the original video is not required.

Without loss of generality, we assume that the last wireless hop is the bottleneck; video data

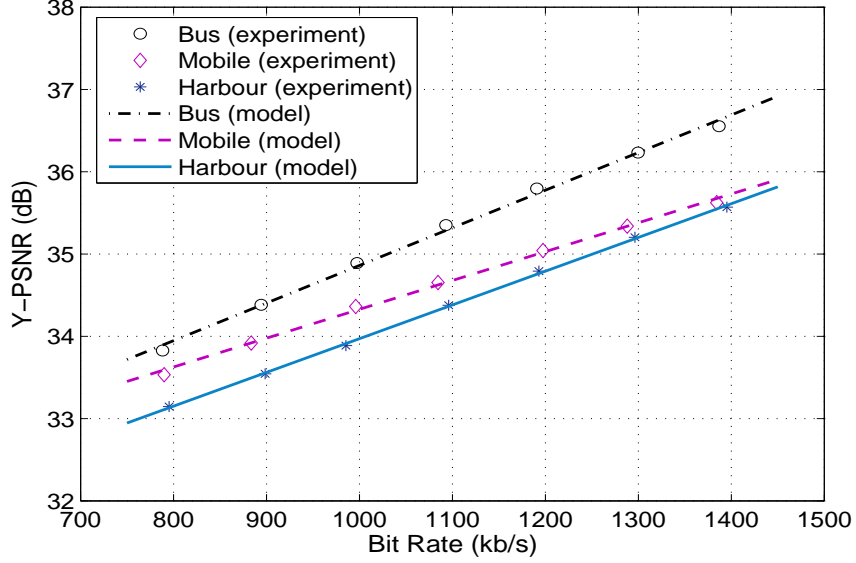


Fig. 6. Rate-distortion curves of three H.264/SVC MGS videos.

is available at the MBS and FBS's when they are scheduled to be transmitted. The quality of reconstructed MGS video can be modeled as [8]:

$$W(R) = \alpha + \beta \times R, \quad (32)$$

where $W(R)$ is the average peak signal-to-noise ratio (PSNR) of the reconstructed video, R is the received data rate, α and β are constants depending on the video sequence and codec.

We verified (32) using an H.264/SVC codec and the *Bus*, *Mobile*, and *Harbour* test sequences. In Fig. 6, the markers are obtained by truncating the encoded video's enhancement layer at different positions to obtain different effective rates, while the curves are computed using (32). The curves fit well with measurements for the three sequences. It is worth noting that PSNR may not be a good measure of video quality as compared with alternative metrics such as MS-SSIM [33]. The main reason for choosing PSNR is that there is a closed-form model relating it to network level metrics–video rate. With the closed-form model, we can have a mathematical formulation of the scheduling/resource allocation problem, and derive effective algorithms. Should such closed-form models be available for MS-SSIM, it is possible to incorporate it into the optimization framework as well.

B. MGS Video over Femtocell CR Networks

In this section, we address the problem of resource allocation for MGS videos over femtocell CR networks. We first examine the case of a single FBS, and then the more general case of multiple non-interfering or interfering FBS's. The algorithms for the single and non-interfering FBS cases are distributed ones and optimal. The algorithm for the interfering FBS case is a centralized one that can be executed at the MBS. To simplify notation, we omit the time slot index t for most of the variables in this Section. For example, x represents a variable for time slot t , x^- represents the variable in time slot $(t - 1)$, and x^+ represents the variable in time slot $(t + 1)$.

1) Case of Single FBS:

a) Formulation: We first consider the case of a single FBS in the CR network, where the FBS can use all the G available channels to stream videos to K active CR users. Let w_j be the PSNR of CR user j at the beginning of time slot t and W_j the PSNR of CR user j at the end of time slot t . In time slot t , w_j is already known; W_j is a random variable that depends on channel condition and primary user activity; and w_j^+ is a *realization* of W_j . Let $\xi_{0,j}$ and $\xi_{1,j}$ indicate the random packet losses from the MBS and FBS, respectively, to CR user j in time slot t . That is, $\xi_{i,j}$ is 1 with probability $\bar{P}_{i,j} = 1 - P_{i,j}$ and 0 with probability $P_{i,j}$. Due to block fading channels, $P_{i,j}$'s do not change within the time slot.

Let $\rho_{0,j}$ and $\rho_{1,j}$ be the portions of time slot t when CR user j receives video data from the MBS and FBS, respectively. The average PSNR is computed every T time slots. We first have $W_j(0) = \alpha_j$, when $t = 0$. In each time slot t , the CR user receives $\xi_{0,j}\rho_{0,j}B_0$ bits through the MBS, and $\xi_{1,j}\rho_{1,j}GB_1$ bits through the FBS (assuming that OFDM is used), which contribute an increase of $\beta(\xi_{0,j}\rho_{0,j}B_0 + \xi_{1,j}\rho_{1,j}GB_1)/T$ to the total PSNR in this T time slot interval, according to (32). Therefore we have the following recursive relationship: $W_j = W_j^- + \beta(\xi_{0,j}\rho_{0,j}B_0 + \xi_{1,j}\rho_{1,j}GB_1)/T = W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}$, where $R_{0,j} = \beta B_0/T$ and $R_{1,j} = \beta B_1/T$.

For proportional fairness, we aim to maximize the sum of the logarithms of the PSNRs of all CR users [34]. We formulate a *multistage stochastic programming problem* by maximizing the

expectation of the logarithm-sum at time T .

$$\begin{aligned}
& \text{maximize: } \sum_{j=1}^K \mathbb{E}[\log(W_j(T))] \\
& \text{subject to: } W_j = W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}, \quad j = 1, \dots, K, \quad t = 1, \dots, T \\
& \quad \sum_{j=1}^K \rho_{i,j} \leq 1, \quad i = 0, 1, \quad t = 1, \dots, T \\
& \quad \rho_{i,j} \geq 0, \quad i = 0, 1, \quad j = 1, \dots, K, \quad t = 1, \dots, T.
\end{aligned} \tag{33}$$

$R_{0,j} = \beta_j B_0/T$ and $R_{1,j} = \beta_j B_1/T$ are constants for the j -th MGS video.

At the beginning of the last time slot T , a realization $\boldsymbol{\xi}_{[T-1]} = [\vec{\xi}_1, \vec{\xi}_2, \dots, \vec{\xi}_{T-1}]$ is known, where $\vec{\xi}_t = [\xi_{0,1}^t, \xi_{0,2}^t, \dots, \xi_{0,K}^t, \xi_{1,1}^t, \dots, \xi_{1,K}^t]$, $t = 1, 2, \dots, T-1$. It can be shown that the multistage stochastic programming problem (33) can be decomposed into T serial sub-problems, each to be solved in a time slot t [7].

$$\begin{aligned}
& \text{maximize: } \sum_{j=1}^K \mathbb{E}\{\log(W_j) | \boldsymbol{\xi}_{[t-1]}\} \\
& \text{subject to: } W_j = W_j^- + \xi_{0,j}\rho_{0,j}R_{0,j} + \xi_{1,j}\rho_{1,j}GR_{1,j}, \quad j = 1, \dots, K \\
& \quad \sum_{j=1}^K \rho_{i,j} \leq 1, \quad i = 0, 1 \\
& \quad \rho_{i,j} \geq 0, \quad i = 0, 1, \quad j = 1, \dots, K,
\end{aligned} \tag{34}$$

where $\mathbb{E}\{\log(W_j) | \boldsymbol{\xi}_{[t-1]}\}$ denotes the *conditional expectation* of $\log(W_j)$ given realization $\boldsymbol{\xi}_{[t-1]}$. W_j^- is known given the realization. When $t = 1$, the conditional expectation becomes an unconditional expectation.

Since a CR user has only one transceiver, it can operate on either one or more licensed channels (i.e., connecting to the FBS) or the common channel (i.e., connecting to the MBS), but not both simultaneously. Assume CR user j operates on the common channel with probability

p_j and one or more licensed channels with probability q_j . We then rewrite problem (34) as

$$\begin{aligned}
& \text{maximize: } \sum_{j=1}^K [p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + q_j \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j})] \\
& \text{subject to: } \sum_{j=1}^K \rho_{i,j} \leq 1, \quad i = 0, 1 \\
& \quad p_j + q_j = 1, \quad j = 1, \dots, K \\
& \quad \rho_{i,j}, p_j, q_j \geq 0, \quad i = 0, 1, j = 1, \dots, K.
\end{aligned}$$

b) Properties: In this section, we analyze the formulated problem (35) and derive its properties. We have Lemmas 1, 2, and 3 and Theorem 1 and provide the proofs in the following.

Lemma 1. *Problem (35) is a convex optimization problem.*

Proof: First, it can be shown that the single term $p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + q_j \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j})$ is a concave function, because its *Hessian matrix* is negative semi-definite. Then, the objective function is concave since the sum of concave functions is also concave. Finally, all the constraints are linear. We conclude that problem (35) is convex with a unique optimal solution. ■

Lemma 2. *If $[\rho, p, q]$ is a feasible solution to problem (35), then $[\rho, q, p]$ is also feasible.*

Proof: Since $[\rho, p, q]$ is feasible, we have $p + q = 1$. Switching the two probabilities, we still have $q + p = 1$. Therefore, the derived new solution is also feasible. ■

Lemma 3. *Let the optimal solution be $[\rho^*, p^*, q^*]$. If $p_j^* \geq q_j^*$, then $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j})$ is greater than or equal to $\bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. And vice versa.*

Proof: Assume $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j})$ is less than $\bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. Since $p_j^* \geq q_j^*$, the sum of the product $p_j^* \bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) + q_j^* \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$ is smaller than the sum of the product $q_j^* \bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) + p_j^* \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$. Thus we can obtain an objective value larger than the optimum by switching the values of p_j^* and q_j^* , which is still feasible according to Lemma 2. This conflicts with the assumption that $[\rho^*, p^*, q^*]$ is optimal. The reverse statement can be proved similarly. ■

Theorem 1. Let the optimal solution be $[\rho^*, p^*, q^*]$. If $p_j^* > q_j^*$, then we have $p_j^* = 1$ and $q_j^* = 0$. Otherwise, we have $p_j^* = 0$ and $q_j^* = 1$.

Proof: If $p_j^* > q_j^*$, we have $\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}^* R_{0,j}) \geq \bar{P}_{1,j} \log(W_j^- + \rho_{1,j}^* G R_{1,j})$ according to Lemma 3. Since the objective function is linear with respect to p_j and q_j , the optimal value can be achieved by setting p_j to its maximum value 1 and q_j to its minimum value 0. The reverse statement can be proved similarly. ■

According to Theorem 1, a CR user is connected to either the MBS or the FBS for the *entire* duration of a time slot in the optimal solution. That is, it does not switch between base stations during a time slot under optimal scheduling.

c) *Distributed Solution Algorithm:* To solve problem (35), we define non-negative *dual variables* $\lambda = [\lambda_0, \lambda_1]$ for the two inequality constraints. The *Lagrangian function* is

$$\begin{aligned} \mathcal{L}(p, \rho, \lambda) &= \sum_{j=1}^K [p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + (1 - p_j) \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j})] + \\ &\quad \lambda_0 (1 - \sum_{j=1}^K \rho_{0,j}) + \lambda_1 (1 - \sum_{j=1}^K \rho_{1,j}) \\ &= \sum_{j=1}^K \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0, \lambda_1) + \lambda_0 + \lambda_1, \end{aligned} \quad (35)$$

where

$$\begin{aligned} \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0, \lambda_1) &= p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + \\ &\quad (1 - p_j) \bar{P}_{1,j} \log(W_j^- + \rho_{1,j} G R_{1,j}) - \lambda_0 \rho_{0,j} - \lambda_1 \rho_{1,j}. \end{aligned}$$

The corresponding problem can be decomposed into K sub-problems and solved iteratively. In Step $\tau \geq 1$, for given $\lambda_0(\tau)$ and $\lambda_1(\tau)$ values, each CR user j solves the following sub-problem using local information.

$$[p_j^*(\tau), \rho_{0,j}^*(\tau), \rho_{1,j}^*(\tau)] = \arg \max_{p_j, \rho_{0,j}, \rho_{1,j} \geq 0} \mathcal{L}_j(p_j, \rho_{0,j}, \rho_{1,j}, \lambda_0(\tau), \lambda_1(\tau)). \quad (36)$$

There is a unique optimal solution since the objective function in (36) is concave. The CR users then exchange their solutions. The *master dual problem*, for given $p(\tau)$ and $\rho(\tau)$, is:

$$\min_{\lambda \geq 0} \mathcal{L}(p(\tau), \rho(\tau), \lambda) = \sum_{j=1}^K \mathcal{L}_j(p_j(\tau), \rho_{0,j}(\tau), \rho_{1,j}(\tau), \lambda_0, \lambda_1) + \lambda_0 + \lambda_1. \quad (37)$$

TABLE III
ALGORITHM FOR THE CASE OF SINGLE FBS

1:	Set $\tau = 0$, $\lambda_0(0)$ and $\lambda_1(0)$ to some nonnegative value;
2:	DO % (each CR user j executes Steps 3–8)
3:	$\rho_{0,j}(\tau) = \left[\frac{\bar{P}_{0,j}}{\lambda_0(\tau)} - \frac{W_j^-}{R_{0,j}} \right]^+$, $\rho_{1,j}(\tau) = \left[\frac{\bar{P}_{1,j}}{\lambda_1(\tau)} - \frac{W_j^-}{R_{1,j}G} \right]^+$;
4:	IF $[\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}(\tau)R_{0,j}) - \lambda_0(\tau)\rho_{0,j}(\tau)] > [\bar{P}_{1,j} \log(W_j^- + \rho_{1,j}(\tau)GR_{1,j}) - \lambda_1(\tau)\rho_{1,j}(\tau)]$
5:	Set $p_j(\tau) = 1$ and $\rho_{1,j}(\tau) = 0$;
6:	ELSE
7:	Set $p_j(\tau) = 0$ and $\rho_{0,j}(\tau) = 0$;
8:	END IF
9:	MBS updates $\lambda_i(\tau + 1)$ as in (38);
10:	$\tau = \tau + 1$;
11:	WHILE $(\sum_{i=0}^1 (\lambda_i(\tau + 1) - \lambda_i(\tau))^2 > \phi)$

Since the Lagrangian function is differentiable, the *gradient iteration* approach can be used.

$$\lambda_i(\tau + 1) = \left[\lambda_i(\tau) - s \times \left(1 - \sum_{j=1}^K \rho_{i,j}^*(\tau) \right) \right]^+, \quad i = 0, 1, \quad (38)$$

where s is a sufficiently small positive *step size* and $[\cdot]^+$ denotes the projection onto the nonnegative axis. The updated $\lambda_i(\tau + 1)$ will again be used to solve the sub-problems, and so forth. Since the problem is convex, we have *strong duality*; the *duality gap* between the primal and dual problems is zero. The dual variables $\lambda(\tau)$ will converge to the optimal values as τ goes to infinity. Since the optimal solution to (36) is unique, the primal variables $p(\tau)$ and $\rho_{i,j}(\tau)$ will also converge to their optimal values when τ is sufficiently large.

The distributed solution procedure is presented in Table III. In the table, Steps 3–8 solve the sub-problem in (36); Step 9 updates the dual variables. The threshold ϕ is a prescribed small value with $0 \leq \phi \ll 1$. The algorithm terminates when the dual variables are sufficiently close to the optimal values.

2) *Case of Multiple Non-interfering FBS's* : We next consider the case of $N > 1$ non-interfering FBS's. The coverages of the FBS's do not overlap with each other, as FBS 1 and 2 in Fig. 5. Consequently, each FBS can use all the available licensed channels without interfering other FBS's. Assume each CR user knows the nearest FBS and is associate with it. Let \mathcal{U}_i denote

the set of CR users associated with FBS i . The resource allocation problem becomes:

$$\begin{aligned}
& \text{maximize: } \sum_{j=1}^K p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + \sum_{i=1}^N \sum_{j \in \mathcal{U}_i} q_j \bar{P}_{i,j} \log(W_j^- + \rho_{i,j} G R_{i,j}) \quad (39) \\
& \text{subject to: } \sum_{j=1}^K \rho_{0,j} \leq 1 \\
& \quad \sum_{j \in \mathcal{U}_i} \rho_{i,j} \leq 1, \quad i = 1, \dots, N \\
& \quad p_j + q_j = 1, \quad j = 1, \dots, K \\
& \quad \rho_{i,j}, p_j, q_j \geq 0, \quad i = 1, \dots, N, j = 1, \dots, K.
\end{aligned}$$

Since all the available channels can be allocated to each FBS with spatial reuse, problem (39) can be solved using the algorithm in Table III with some modified notation: $\rho_{1,j}(\tau)$ now becomes $\rho_{i,j}(\tau)$ and $\lambda_1(\tau)$ becomes $\lambda_i(\tau)$, $i = 1, \dots, N$. The dual variables are iteratively updated as

$$\lambda_0(\tau + 1) = \left[\lambda_0(\tau) - s \times \left(1 - \sum_{j=1}^K \rho_{0,j}^*(\tau) \right) \right]^+ \quad (40)$$

$$\lambda_i(\tau + 1) = \left[\lambda_i(\tau) - s \times \left(1 - \sum_{j \in \mathcal{U}_i} \rho_{i,j}^*(\tau) \right) \right]^+, \quad i = 1, \dots, N. \quad (41)$$

The modified solution algorithm is presented in Table IV. As in the case of single FBS, the algorithm is jointly executed by the CR users and MBS, by iteratively updating the dual variables $\lambda_0(\tau)$ and $\lambda_i(\tau)$'s, and the resource allocations $\rho_{0,j}^*(\tau)$ and $\rho_{i,j}^*(\tau)$'s. It can be shown that the distributed algorithm can produce the optimal solution for problem (39).

3) Case of Multiple Interfering FBS's:

a) *Formulation:* Finally, we consider the case of multiple interfering FBS's. Assume that the coverages of some FBS's overlap with each other, as FBS 3 and 4 in Fig. 5. They cannot use the same channel simultaneously, but have to compete for the available channels in the transmission phase. Define *channel allocation variables* $c_{i,m}$ for time slot t as:

$$c_{i,m} = \begin{cases} 1, & \text{if channel } m \text{ is allocated to FBS } i \\ 0, & \text{otherwise.} \end{cases} \quad (42)$$

Given an allocation, the expected number of available channels for FBS i is $G_i = \sum_{m \in \mathcal{A}(t)} c_{i,m} P_m^A$.

We use *interference graph* to model the case of overlapping coverages, which is defined below.

TABLE IV
ALGORITHM FOR THE CASE OF MULTIPLE NON-INTERFERING FBS'S

1:	Set $\tau = 0$, and $\lambda_0(0)$ and $\lambda_i(0)$ to some nonnegative values, for all i ;
2:	DO % (each CR user j executes Steps 3–8)
3:	$\rho_{0,j}(\tau) = \left[\frac{\bar{P}_{0,j}}{\lambda_0(\tau)} - \frac{W_j^-}{R_{0,j}} \right]^+$, $\rho_{i,j}(\tau) = \left[\frac{\bar{P}_{i,j}}{\lambda_i(\tau)} - \frac{W_j^-}{R_{i,j}G} \right]^+$;
4:	IF $[\bar{P}_{0,j} \log(W_j^- + \rho_{0,j}(\tau)R_{0,j}) - \lambda_0(\tau)\rho_{0,j}(\tau)] > [\bar{P}_{i,j} \log(W_j^- + \rho_{i,j}(\tau)GR_{i,j}) - \lambda_i(\tau)\rho_{i,j}(\tau)]$
5:	Set $p_j(\tau) = 1$ and $\rho_{i,j}(\tau) = 0$;
6:	ELSE
7:	Set $p_j(\tau) = 0$ and $\rho_{0,j}(\tau) = 0$;
8:	END IF
9:	MBS updates $\lambda_i(\tau + 1)$ as in (40) and (41);
10:	$\tau = \tau + 1$;
11:	WHILE $\left(\sum_{i=0}^N (\lambda_i(\tau + 1) - \lambda_i(\tau))^2 > \phi \right)$



Fig. 7. Interference graph for the femtocell CR network shown in Fig. 5.

Definition 1. An interference graph $G_I = (V_I, E_I)$ is an undirected graph where each vertex represents an FBS and each edge indicates interference between the two end FBS's.

For the example given in Fig. 5, we can derive an interference graph as shown in Fig. 7. FBS 3 and 4 cannot use the same channel simultaneously, as summarized in the following lemma.

Lemma 4. If channel m is allocated to FBS i , the neighboring vertices of FBS i in the interference graph G_I , denoted as $\mathcal{R}(i)$, cannot use the same channel m simultaneously.

Further define index variables d_i^k as

$$d_i^k = \begin{cases} 1, & \text{if FBS } i \text{ is an endpoint of link } k \in G_I \\ 0, & \text{otherwise.} \end{cases} \quad (43)$$

The interference constraint can be described as $\sum_{i=1}^N d_i^k c_{i,m} \leq 1$, for $m = 0, \dots, M$, and for all

link $k \in G_I$. We then have the following problem formulation.

$$\begin{aligned}
& \text{maximize: } \sum_{j=1}^K p_j \bar{P}_{0,j} \log(W_j^- + \rho_{0,j} R_{0,j}) + \sum_{i=1}^N \sum_{j \in \mathcal{U}_i} q_j \bar{P}_{i,j} \log(W_j^- + \rho_{i,j} G_i R_{i,j}) \quad (44) \\
& \text{subject to: } \sum_{j=1}^K \rho_{0,j} \leq 1 \\
& \quad \sum_{j \in \mathcal{U}_i} \rho_{i,j} \leq 1, \quad i = 1, \dots, N \\
& \quad p_j + q_j = 1, \quad j = 1, \dots, K \\
& \quad G_i = \sum_{m \in \mathcal{A}(t)} c_{i,m} P_m^A, \quad i = 1, \dots, N \\
& \quad \sum_{i=1}^N d_i^k c_{i,m} \leq 1, m = 0, \dots, M, \text{ for link } k \in G_I, \\
& \quad \rho_{i,j}, p_j, q_j, c_{i,m} \geq 0, \quad i = 1, \dots, N, \quad j = 1, \dots, K, \quad m = 0, \dots, M.
\end{aligned}$$

b) Solution Algorithm : The optimal solution to problem (44) depends on the channel allocation variables $c_{i,m}$. Problem (44) can be solved with the algorithm in Table IV if the $c_{i,m}$'s are known. Let $Q(\mathbf{c})$ be the suboptimal objective value for a given channel allocation \mathbf{c} , where $\mathbf{c} = [\vec{c}_1, \vec{c}_2, \dots, \vec{c}_N]$ and \vec{c}_i is a vector of elements $c_{i,m}$, for FBS i and channels $m \in \mathcal{A}(t)$. If all the FBS's are disjointedly distributed with no overlap, each FBS can use all the available channels. We have $c_{i,m} = 1$ for all i and $m \in \mathcal{A}(t)$, i.e., it is reduced to the case in Section IV-B2.

To solve problem (44), we first apply a *greedy algorithm* to allocate the available channels in $\mathcal{A}(t)$ to the FBS's (i.e., to determine \mathbf{c}). We then apply the algorithm in Table IV with the computed \mathbf{c} to obtain a near-optimal solution. Let $e_{i,m}$ be a matrix with 1 at position $\{i, m\}$ and 0 at all other positions, representing the allocation of channel $m \in \mathcal{A}(t)$ to FBS i . The greedy channel allocation algorithm is given in Table V, where the FBS-channel pair that can achieve the largest increase in $Q(\cdot)$ is chosen in each iteration. The worst case complexity of the greedy algorithm is $O(N^2 M^2)$.

c) Performance Lower Bound: We next present a lower bound for the greedy algorithm. Let $e(l)$ be the l -th FBS-channel pair chosen in the greedy algorithm, and π_l denote the sequence $\{e(1), e(2), \dots, e(l)\}$. The increase in object value (44) due to the l -th allocated FBS-channel pair is denoted as

$$\Delta_l := \Delta(\pi_l, \pi_{l-1}) = Q(\pi_l) - Q(\pi_{l-1}). \quad (45)$$

TABLE V
CHANNEL ALLOCATION ALGORITHM FOR CASE OF INTERFERING FBS'S

1:	Initialize \mathbf{c} to a zero matrix, FBS set $\mathcal{N} = \{1, \dots, N\}$, and FBS-channel set $\mathcal{C} = \mathcal{N} \times \mathcal{A}(t)$;
2:	WHILE (\mathcal{C} is not empty)
3:	Find FBS-channel pair $\{i', m'\}$, such that $\{i', m'\} = \arg \max_{\{i, m\} \in \mathcal{C}} \{Q(\mathbf{c} + \mathbf{e}_{i, m}) - Q(\mathbf{c})\};$
4:	Set $\mathbf{c} = \mathbf{c} + \mathbf{e}_{i', m'}$;
5:	Remove $\{i', m'\}$ from \mathcal{C} ;
6:	Remove $\mathcal{R}(i') \times m'$ from \mathcal{C} ;
7:	END WHILE

Since $Q(\pi_0) = Q(\emptyset) = 0$, we have

$$\begin{aligned} \sum_{l=1}^L \Delta_l &= Q(\pi_L) - Q(\pi_{L-1}) + \dots + Q(\pi_1) - Q(\pi_0) \\ &= Q(\pi_L) - Q(\pi_0) = Q(\pi_L). \end{aligned}$$

For two FBS-channel pairs $e(l)$ and $e(l')$, we say $e(l)$ *conflicts with* $e(l')$ when there is an edge connecting the FBS in $e(l)$ and the FBS in $e(l')$ in the interference graph G_I , and the two FBS's choose the same channel. Let Ω be the global optimal solution. We define ω_l as the subset of Ω that conflicts with allocation $e(l)$ but not with the previous allocations $\{e(1), e(2), \dots, e(l-1)\}$.

Lemma 5. *Assume the greedy algorithm in Table V stops in L steps. The global optimal solution Ω can be partitioned into L non-overlapping subsets ω_l , $l = 1, 2, \dots, L$.*

Proof: According to the definition of ω_l , the L subsets of the optimal solution Ω do not intersect with each other. Assume the statement is false, then the union of these L subsets is not equal to the optimal set Ω . Let the *set difference* be $\omega_{L+1} = \Omega \setminus (\cup_{l=1}^L \omega_l)$. By definition, ω_{L+1} does not conflict with the existing L allocations $\{e(1), \dots, e(L)\}$, meaning that the greedy algorithm can continue to at least the $(L+1)$ -th step. This conflicts with the assumption that the greedy algorithm stops in L steps. It follows that $\Omega = \cup_{l=1}^L \omega_l$. ■

Let $\Delta(\pi_2, \pi_1) = Q(\pi_2) - Q(\pi_1)$ denote the difference between two feasible allocations π_1 and π_2 . We next derive a lower bound on the performance of the greedy algorithm. We assume two properties for function $\Delta(\pi_2, \pi_1)$ in the following.

Property 1. Consider FBS-channel pair sets π_1 , π_2 , and σ , satisfying $\pi_1 \subseteq \pi_2$ and $\sigma \cap \pi_2 = \emptyset$. We have $\Delta(\pi_2 \cup \sigma, \pi_1 \cup \sigma) \leq \Delta(\pi_2, \pi_1)$.

Property 2. Consider FBS-channel pair sets π , σ_1 , and σ_2 satisfying $\sigma_1 \cap \pi = \emptyset$, $\sigma_2 \cap \pi = \emptyset$, and $\sigma_1 \cap \sigma_2 = \emptyset$. We have $\Delta(\sigma_1 \cup \sigma_2 \cup \pi, \pi) \leq \Delta(\sigma_1 \cup \pi, \pi) + \Delta(\sigma_2 \cup \pi, \pi)$.

In Property 1, we have $\sigma \cap \pi_1 = \emptyset$ since $\pi_1 \subseteq \pi_2$ and $\sigma \cap \pi_2 = \emptyset$. This property states that the incremental objective value does not get larger as more channels are allocated and as the objective value gets larger. Property 2 states that the incremental objective value achieved by allocating multiple FBS-channel pair sets does not exceed the sum of the incremental objective values achieved by allocating each individual FBS-channel pair set. These are generally true for many resource allocation problems [34].

Since we choose the maximum incremental allocation in each step of the greedy algorithm, we have Lemma 6 that directly follows Step 3 in Table V.

Lemma 6. For any FBS-channel pair $\sigma \in \omega_l$, we have $Q(\pi_{l-1} \cup \sigma) - Q(\pi_{l-1}) = \Delta(\pi_{l-1} \cup \sigma, \pi_{l-1}) \leq \Delta_l$.

Lemma 7. Assume the greedy algorithm stops in L steps, we have

$$Q(\Omega) \leq Q(\pi_L) + \sum_{l=1}^L \sum_{\sigma \in \omega_l} \Delta(\sigma \cup \pi_{l-1}, \pi_{l-1}).$$

Proof: The following inequalities hold true according to the properties of the $\Delta(\cdot, \cdot)$ function:

$$\begin{aligned} Q((\cup_{i=l+1}^L \omega_i) \cup \pi_l) &= Q((\cup_{i=l+2}^L \omega_i) \cup \pi_l) + \Delta((\cup_{i=l+1}^L \omega_i) \cup \pi_l, (\cup_{i=l+2}^L \omega_i) \cup \pi_l) \\ &\leq Q((\cup_{i=l+2}^L \omega_i) \cup \pi_l) + \Delta(\omega_{l+1} \cup \pi_l, \pi_l) \\ &\leq Q((\cup_{i=l+2}^L \omega_i) \cup \pi_{l+1}) + \Delta(\omega_{l+1} \cup \pi_l, \pi_l) \\ &\leq Q((\cup_{i=l+2}^L \omega_i) \cup \pi_{l+1}) + \sum_{\sigma \in \omega_{l+1}} \Delta(\sigma \cup \pi_l, \pi_l). \end{aligned}$$

We have $\pi_0 = \emptyset$ and $\omega_{L+1} = \emptyset$ (see Lemma 5). With induction from $l = 0$ to $l = L - 1$, we have $Q((\cup_{i=1}^L \omega_i) \cup \emptyset) = Q(\Omega)$ and $Q(\Omega) \leq Q(\pi_L) + \sum_{l=1}^L \sum_{\sigma \in \omega_l} \Delta(\sigma \cup \pi_{l-1}, \pi_{l-1})$. ■

Lemma 8. The maximum size of ω_l is equal to the degree, in the interference graph G_I , of the FBS selected in the l -th step of the greedy algorithm, which is denoted as $D(l)$.

Proof: Once FBS i is allocated with channel m , the neighboring FBS's in G_I , $\mathcal{R}(i)$, cannot use the same channel m anymore due to the interference constraint. The maximum number of FBS-channel pairs that conflict with the selected FBS-channel pair $\{i, m\}$, i.e., the maximum size of ω_l , is equal to the degree of FBS i in G_I . ■

Then we have Theorem 2 that provides a lower bound on the objective value achieved by the greedy algorithm given in Table V.

Theorem 2. *The greedy algorithm can achieve an objective value that is at least $\frac{1}{1+D_{max}}$ of the global optimum, where D_{max} is the maximum node degree in the interference graph G_I of the femtocell CR network.*

Proof: According to Lemmas 7 and 8, we have:

$$\begin{aligned} Q(\Omega) &\leq Q(\pi_L) + \sum_{l=1}^L D(l)\Delta_l = Q(\pi_L) + \bar{D} \sum_{l=1}^L \Delta_l \\ &= (1 + \bar{D})Q(\pi_L), \end{aligned} \quad (46)$$

where $\bar{D} = \sum_{l=1}^L D(l)\Delta_l / \sum_{l=1}^L \Delta_l$. The second equality is due to the facts that $\sum_{l=1}^L \Delta_l = Q(\pi_L)$.

To further simplify the bound, we replace $D(l)$ with the maximum node degree D_{max} . We then have $\bar{D} \leq \sum_{l=1}^L D_{max}\Delta_l / \sum_{l=1}^L \Delta_l = D_{max}$ and

$$\frac{1}{1 + D_{max}}Q(\Omega) \leq Q(\pi_L) \leq Q(\Omega), \quad (47)$$

which provides a lower bound on the performance of the greedy algorithm. ■

When there is a single FBS in the CR network, we have $D_{max} = 0$ and $Q(\pi_L) = Q(\Omega)$ according to Theorem 2. The proposed algorithm produces the optimal solution. In the case of multiple non-interfering FBS's, we still have $D_{max} = 0$ and can obtain the optimal solution using the proposed algorithm. For the femtocell CR network given in Fig. 5 (with interference graph shown in Fig. 7), we have $D_{max} = 1$ and the low bound is a half of the global optimal. Note that (46) provides a tighter bound for the optimum than (47), but with higher complexity. These are interesting performance bounds since they bound the achievable video quality, an application layer performance measure, rather than lower layer metrics (e.g., bandwidth or time share).

C. Simulation Results

We evaluate the performance of the proposed algorithms using MATLAB and JSVM 9.13 Video codec. Two scenarios are used in the simulations: a single FBS CR network and a CR network with interfering FBS's. In every simulation, we compare the proposed algorithms with the following three more straightforward heuristic schemes:

- Heuristic 1 based on *equal allocation*: each CR user chooses the better channel (i.e., the common channel or a licensed channel) based on the channel conditions; time slots are equally allocated among active CR users;
- Heuristic 2 exploiting *multiuser diversity*: the MBS and each FBS chooses one active CR user with the best channel condition; the entire time slot is allocated to the selected CR user.
- *SCA-MAC* proposed in [35]: with this scheme, the successful transmission rate is evaluated based on channel packet loss rate and collision probability with primary users; the channel-user pair with the highest transmission probability is selected.

We choose SCA-MAC because it adopts similar models and assumptions as in this paper. Once the channels are selected, the same distributed algorithm is used for scheduling video data for all the three schemes.

We adopt the Raleigh block fading model and the packet loss probability is between [0.004, 0.028]. The frame rate is set to 30 fps and the GoP size is 16. The base layer mode is set to be AVC compatible. The motion search mode is set to Fast Search with search range 32. Each point in the figures presented in this section is the average of 10 simulation runs with different random seeds. We plot 95% confidence intervals in the figures, which are generally negligible.

1) *Case of Single FBS*: In the first scenario, there are $M = 8$ channels and the channel parameters P_{01}^m and P_{10}^m are set to 0.4 and 0.3, respectively, for all m . The maximum allowable collision probability γ_m is set to 0.2 for all m . There is one FBS and three active CR users. Three Common Intermediate Format (CIF, 352×288) video sequences are streamed to the CR users, i.e., *Bus* to CR user 1, *Mobile* to CR user 2, and *Harbor* to CR user 3. We have $T = 10$ as the delivery deadline. Both probabilities of false alarm ϵ and miss detection δ are set to 0.3 for all the FBS's and CR users, unless otherwise specified.

First we investigate the convergence of the distributed algorithm. The traces of the two dual

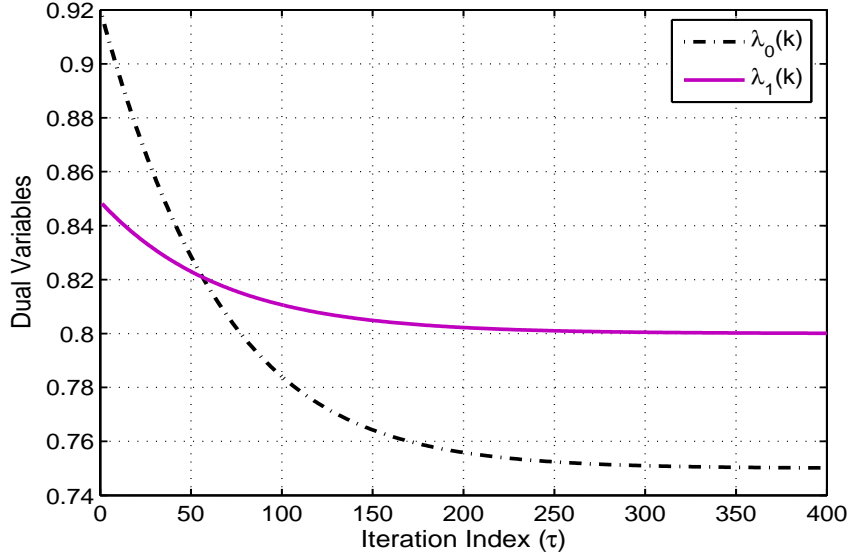


Fig. 8. Convergence of the two dual variables in the single FBS case.

variables are plotted in Fig. 8. To improve the convergence speed, the correlation in adjacent time slots can be exploited. In particular, we set the optimal values for the optimization variables in the previous time slot as the initialization values for the variables in the current time slot. By doing so, the convergence speed can be improved. It can be seen that both dual variables converge to their optimal values after 300 iterations. After convergence, the optimal solution for the primary problem can be obtained.

Our proposed scheme achieves the best performance among the three algorithms, with up to 4.3 dB improvement over the two heuristic schemes and up to 2.5 dB over SCA-MAC. Such gains are significant with regard to video quality, since a 0.5 dB difference is distinguishable by human eyes. Compared to the two heuristic schemes and SCA-MAC, the video quality of our proposed scheme is well balanced among the three users, indicating better fairness performance.

In Fig. 9, we examine the impact of the number of channels M on received video quality. First, we validate the video quality measure used in our formulation by comparing the PSNR value computed using (32) with that computed from real decoded video frames. The average PSNR for three received videos are plotted in the figure. It can be seen that the real PSNRs are very close to those predicted by (32), with overlapping confidence intervals. This is also consistent with the results shown in Fig. 6. Second, as expected, the more licensed channels, the

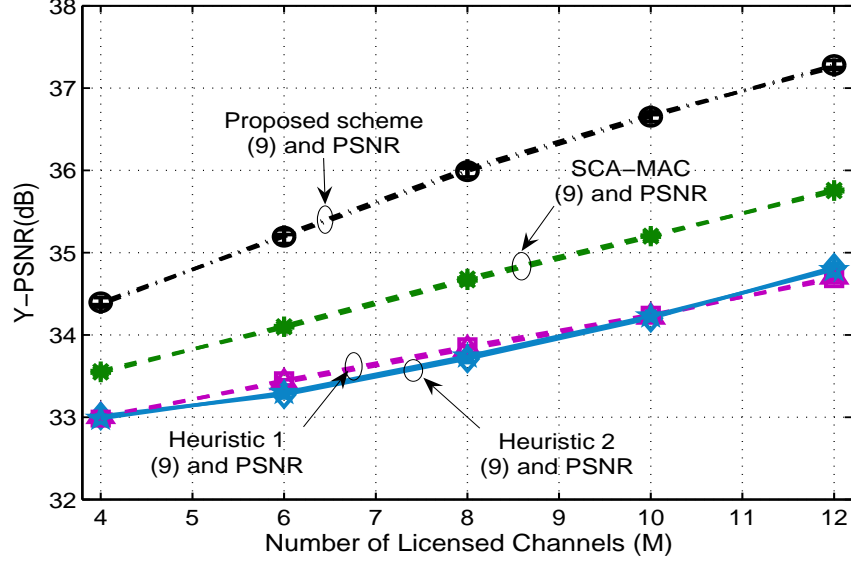


Fig. 9. Single FBS: received video quality vs. number of channels (computed with (9) and measured by PSNR).

more spectrum opportunities for CR users and the higher PSNR for received videos. SCA-MAC performs better than two heuristics, but is inferior to the proposed scheme.

We also plot the MS-SSIM of the received videos at the three CR users in Fig. 10 [33]. Similar observations can be made from the MS-SSIM plot. All MS-SSIMs for the four curves are more than 0.97 and very close to 1. The proposed scheme still outperforms the other three schemes. In the remaining figures, we will use model predicted PSNR values, since the model (32) is sufficient to predict the real video quality.

In Fig. 11, we demonstrate the impact of channel utilization η on received video quality. The average PSNRs achieved by the four schemes are plotted when η is increased from 0.3 to 0.7. Intuitively, a smaller η allows more spectrum opportunities for video transmission. This is illustrated in the figure where all the three curves decrease as η gets larger. The performance of both heuristics are close and the proposed scheme achieves a gain about 3 dB over the heuristics and 2 dB over SCA-MAC.

We also compare the MGS and FGS videos while keeping other parameters identical. We find that MGS video achieves over 0.5 dB gain in video quality over FGS video. The results are omitted for brevity.

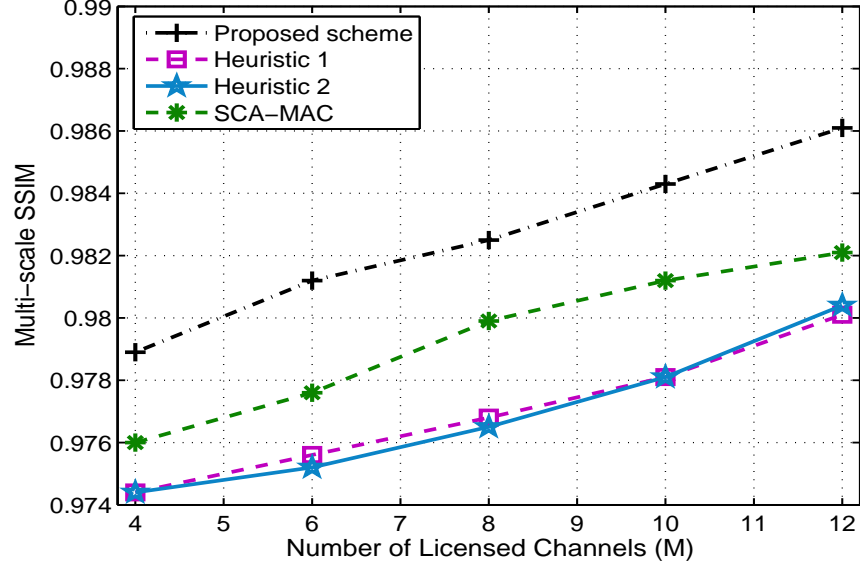


Fig. 10. Single FBS: received video quality vs. number of channels (measured by MS-SSIM).

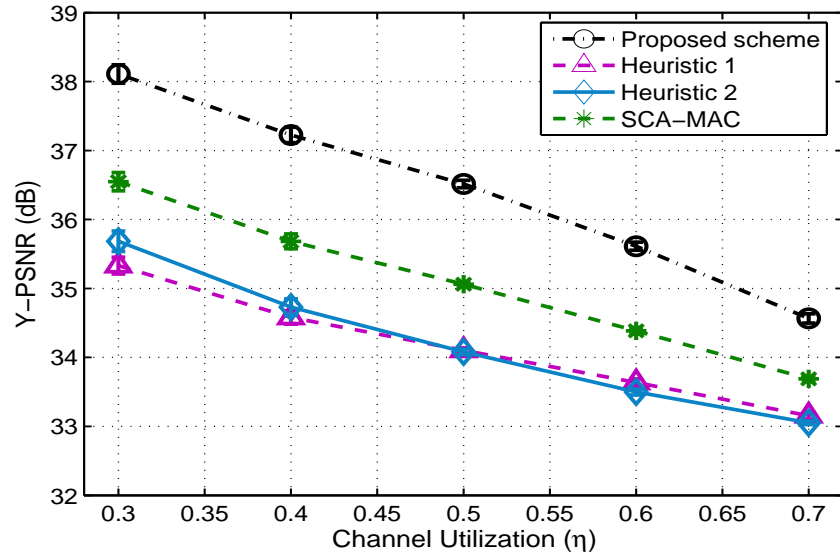


Fig. 11. Single FBS: received video quality vs. channel utilization.

2) *Case of Interfering FBS's*: We next investigate the second scenario with three FBS's, and each FBS has three active CR users. Each FBS streams three different videos to the corresponding CR users. The coverages of FBS 1 and 2 overlap with each other, and the coverages of FBS 2 and 3 overlap with each other.

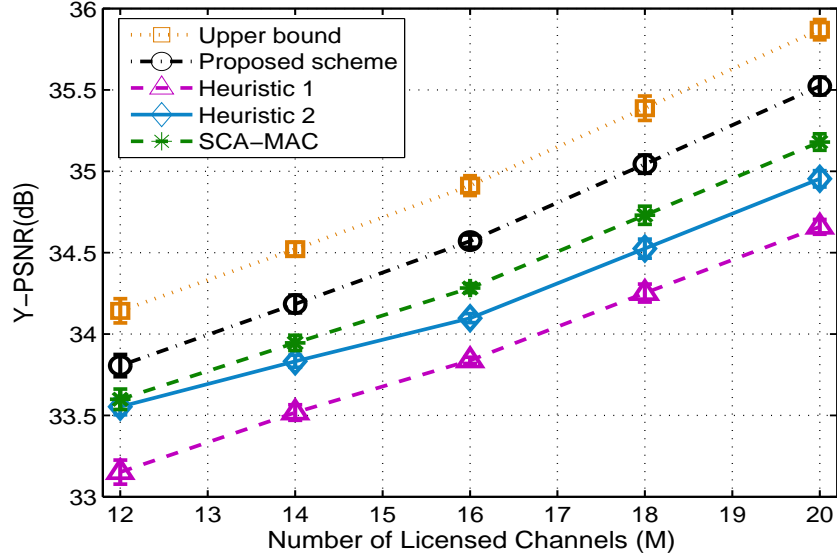


Fig. 12. Interfering FBS's: received video quality vs. number of channels.

In Fig. 12, we examine the impact of the number of channels M on the received video quality. The average PSNRs of all the active CR users are plotted in the figure when we increase M from 12 to 20 with step size 2. As mentioned before, more channels imply more transmission opportunities for video transmission. In this scenario, heuristic 2 (with a multiuser diversity approach) outperforms heuristic 1 (with an equal allocation approach). But its PSNRs are still about 0.3 ~ 0.5 dB lower than those of the proposed algorithm. The proposed scheme has up to 0.4 dB improvement over SCA-MAC. In Fig. 12, we also plot an upper bound on the optimal objective value, which is obtained as in (46). It can be seen that the performance of our proposed scheme is close to optimal solution since the gap between the upper bound and our scheme is generally small (about 0.5 dB).

Next, we examine the impact of sensing errors on the received video quality. In Fig. 13, we test five pairs of $\{\epsilon, \delta\}$ values: $\{0.2, 0.48\}$, $\{0.24, 0.38\}$, $\{0.3, 0.3\}$, $\{0.38, 0.24\}$, and $\{0.48, 0.2\}$. It is interesting to see that the performance of all the four schemes get worse when the probability of one of the two sensing errors gets large. We can trade-off between false alarm and miss detection probabilities to find the optimal operating point for the spectrum sensors. Moreover, the dynamic range of video quality is not big for the range of sensing errors simulated, compared to that in Fig. 12. This is because both sensing errors are modeled and treated in the algorithms. Again,

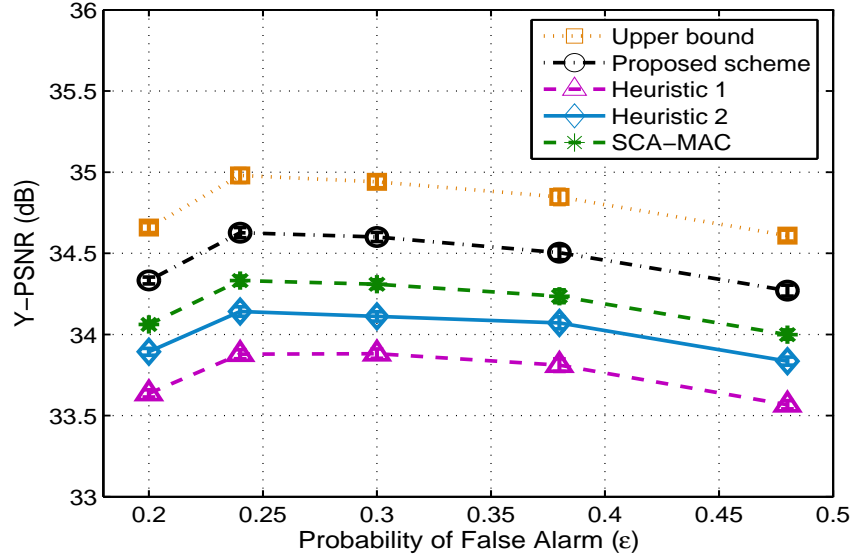


Fig. 13. Interfering FBS's: received video quality vs. sensing error probability.

our proposed scheme outperforms the two heuristic schemes and SCA-MAC with considerable margins for the entire range.

We also investigate the impact of the bandwidth of the common channel B_0 . In this simulation, we fix B_1 at 0.3 Mbps and increase B_0 from 0.1 Mbps to 0.5 Mbps with step size 0.1 Mbps. The results are presented in Fig. 14. We notice that the average video quality increases rapidly as the common channel bandwidth is increased from 0.1 Mbps to 0.3 Mbps. Beyond 0.3 Mbps, the increases of the PSNR curves slow down and the curves get flat. This implies that a very large bandwidth for the common channel is not necessary, since the gain for additional bandwidth diminishes as B_0 gets large. Again, the proposed scheme outperforms the other three schemes and the gap between our scheme and the upper bound is small.

Next, we stop the distributed algorithm after a fixed amount of time, and evaluate the suboptimal solutions. In particular, we vary the duration of time slots, and let the distributed algorithm run for 5% of the time slot duration at the beginning of the time slot. Then the solution obtained this way will be used for the video data transmissions. The results are presented in Fig. 15. It can be seen that when the time slot is 5 ms, the algorithm does not converge after $5\% \times 5 = 0.25$ ms and the PSNR produced by the distributed algorithm is close to that of Heuristic 1, and lower than those of Heuristic 2 and SCA-MAC. When the time slot is sufficiently large,

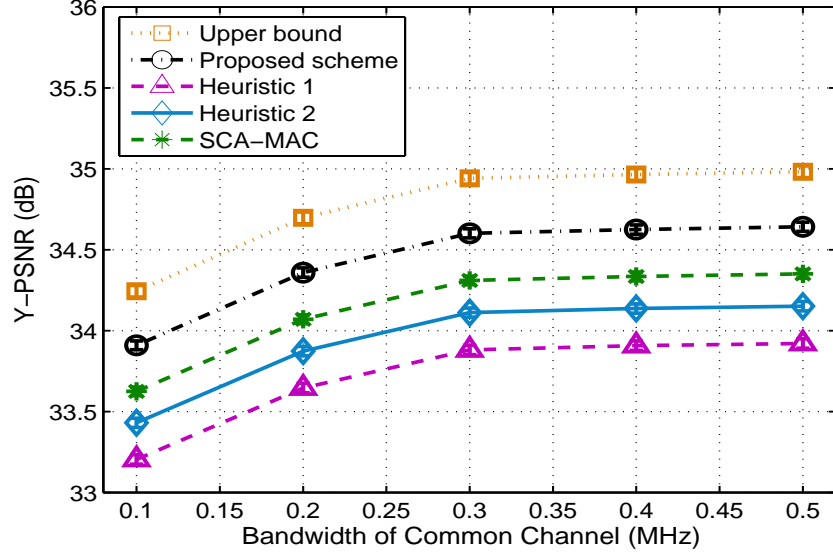


Fig. 14. Interfering FBS's: received video quality vs. bandwidth of the common channel.

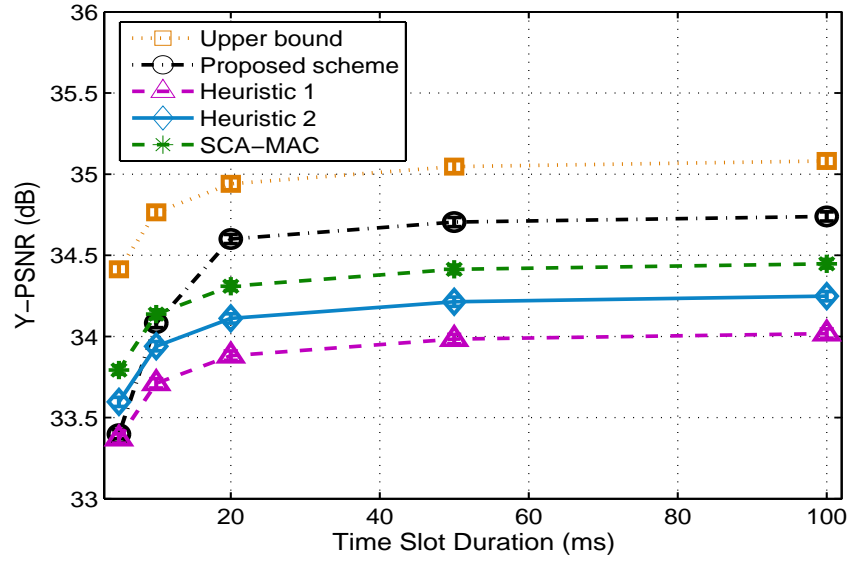


Fig. 15. Video quality achieved by the algorithms when they are only executed for 5% of the time slot duration.

the algorithm can get closer to the optimal and the proposed algorithm produces better video quality as compared to the two heuristic algorithms and SCA-MAC. Beyond 20 ms, the increase in PSNR is small since all the curves get flat. Therefore the proposed algorithm could be useful even when there is no time for it to fully converge to the optimal.

During the simulations, we find the collision rate with primary users are strictly kept below the prescribed collision tolerance γ . These results are omitted for brevity.

V. CONCLUSIONS

In this paper, we first investigated data multicast in femtocell networks consisting of an MBS and multiple FBS's. We adopted SC and SIC for multicast data and investigated how to assign transmit powers for the packet levels. The objective was to minimize the total BS power consumption, while guaranteeing successful decoding of the multicast data at each user. We developed optimal and near-optimal algorithms with low computational complexity, as well as performance bounds. The algorithms were evaluated with simulations and are shown to outperform a heuristic with considerable gains.

Next, we investigated the problem of streaming multiple MGS videos in a femtocell CR network. We formulated a multistage stochastic programming problem considering various design factors across multiple layers. We developed a distributed algorithm that can produce optimal solutions in the case of non-interfering FBS's, and a greedy algorithm for near-optimal solutions in the case of interfering FBS's with a proved lower bound. The proposed algorithms are evaluated with simulations and are shown to outperform three alternative schemes with considerable gains.

REFERENCES

- [1] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [2] R. Kim, J. S. Kwak, and K. Etemad, "WiMAX femtocell: requirements, challenges, and solutions," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 84–91, Sept. 2009.
- [3] I. Guvenc, S. Saunders, O. Oyman, H. Claussen, and A. Gatherer, "Femtocell networks," *EURASIP J. Wireless Comm. and Networking*, 2010, article ID 367878, 2 pages, doi:10.1155/2010/367878.
- [4] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2006.
- [5] L. Li, R. Alimi, R. Ramjee, H. Viswanathan, and Y. R. Yang, "muNet: Harnessing multiuser capacity in wireless mesh networks," in *Proc. IEEE INFOCOM'09 Mini Symp.*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2876–2880.
- [6] D. Hu, S. Mao, Y. Hou, and J. Reed, "Scalable video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 334–344, Apr. 2010.
- [7] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3501–3511, Nov. 2010.
- [8] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194–1203, Sept. 2007.

- [9] V. Chandrasekhar, J. G. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316–4328, Aug. 2009.
- [10] H.-C. Lee, D.-C. Oh, and Y.-H. Lee, "Mitigation of inter-femtocell interference with adaptive fractional frequency reuse," in *Proc. IEEE ICC'10*, Cape Town, South Africa, May 2010, pp. 1–5.
- [11] F.-S. Chu and K.-C. Chen, "Mitigation of macro-femto co-channel interference by spatial channel separation," in *Proc. IEEE VTC-Spring'11*, Budapest, Hungary, May 2011, pp. 1–5.
- [12] S. Rangan, "Femto-macro cellular interference control with subband scheduling and interference cancelation," in *Proc. IEEE GLOBECOM'10 Workshops*, Miami, FL, Dec. 2010, pp. 695–700.
- [13] Z. Bharucha, H. Haas, G. Auer, and I. Cosovic, "Femto-cell resource partitioning," in *Proc. IEEE GLOBECOM Workshops*, Honolulu, HI, Dec. 2009, pp. 1–6.
- [14] R. Madan, A. Sampath, A. Khandekar, J. Borran, and N. Bhushan, "Distributed interference management and scheduling in LTE-A femto networks," in *Proc. IEEE GLOBECOM'10*, Miami, FL, Dec. 2010, pp. 1–5.
- [15] S.-M. Cheng, S.-Y. Lien, F.-S. Chu, and K.-C. Chen, "On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 40–47, 2011.
- [16] S. Kaimaletu, R. Krishnan, S. Kalyani, N. Akhtar, and B. Ramamurthi, "Cognitive interference management in heterogeneous femto-macro cell networks," in *Proc. IEEE ICC'11*, Kyoto, Japan, June 2011, pp. 1–6.
- [17] D. Hu and S. Mao, "Multicast in femtocell networks: a successive interference cancellation approach," in *Proc. IEEE GLOBECOM'11*, Houston, TX, Dec. 2011, pp. 1–6.
- [18] S. Sen, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Successive interference cancellation: A back-of-the-envelope perspective," in *Proc. ACM Hotnets'10*, Monterey, CA, Oct. 2010, pp. 1–6.
- [19] C. A. S. Jean and B. Jabbari, "On game-theoretic power control under successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1655–1657, Apr. 2009.
- [20] C. S. Park and K. B. Lee, "Transmit power allocation for successive interference cancellation in multicode MIMO systems," *IEEE Trans. Commun.*, vol. 56, no. 12, pp. 2200–2213, Dec. 2008.
- [21] N. Benvenuto, G. Carnevale, and S. Tomasin, "Joint power control and receiver optimization of CDMA transceivers using successive interference cancellation," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 563–573, Mar. 2007.
- [22] A. Agrawal, J. G. Andrews, J. M. Cioffi, and T. Meng, "Iterative power control for imperfect successive interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 878–884, May 2005.
- [23] J. G. Andrews and T. H. Meng, "Optimum power control for successive interference cancellation with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 375–383, Mar. 2003.
- [24] H.-P. Shiang and M. van der Schaar, "Dynamic channel selection for multi-user video streaming over cognitive radio networks," in *Proc. IEEE ICIP'08*, San Diego, CA, Oct. 2008, pp. 2316–2319.
- [25] L. Ding, S. Pudlewski, T. Melodia, S. Batalama, J. Matyjas, and M. Medley, "Distributed spectrum sharing for video streaming in cognitive radio ad hoc networks," in *Intl. Workshop on Cross-layer Design in Wireless Mobile Ad Hoc Networks*, Niagara Falls, Canada, Sept. 2009, pp. 1–13.
- [26] D. Hu and S. Mao, "On medium grain scalable video streaming over femtocell cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 641–651, Apr. 2012.
- [27] H. Luo, S. Ci, and D. Wu, "A cross-layer design for the performance improvement of real-time video transmission of secondary users over cognitive radio networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1040–1048, Aug. 2011.

- [28] D. Hu and S. Mao, "Resource allocation for medium grain scalable videos over femtocell cognitive radio networks," in *Proc. IEEE ICDCS'11*, Minneapolis, MN, June 2011, pp. 258–267.
- [29] U. Ewaldsson, "Cut your network's electricity bill and carbon footprint," *Global Telecoms Business*, Feb. 28, 2010.
- [30] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.
- [31] T. Rappaport, *Wireless Communications: Principles & Practice*, 2nd ed. Indianapolis, IN: Prentice Hall PTR, 2001.
- [32] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.
- [33] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment using structural distortion measurement," *Signal Processing: Image Commun.*, no. 2, pp. 121–132, Feb. 2004.
- [34] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, no. 3, pp. 237–252, Mar. 1998.
- [35] A. Hsu, D. Wei, and C. Kuo, "A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks," in *Proc. IEEE WCNC'07*, Hong Kong, P.R. China, Mar. 2007, pp. 105–110.